Introduction

From the Study of Media Upheavals
to an Upheaval in Media Studies

Looking at the history of media in the past 150 years or so, we can observe a long succession of developments and refinements on the technological level as well as on the levels of production and reception, including cultural adaptations and social acceptance. At some points in time, these developments seem to gain momentum and bring about a change so radical that it reflects on society at large. At the research centre "Media Upheavals", based at the University of Siegen and funded by Deutsche Forschungsgemeinschaft, we take into view two periods that stand out for their far-reaching technological change and the impact they have had on society: At the beginning of the 20th century, a radical change was perceived and commonly pinpointed on film and other media for recording and reproducing 'life'. Today, at the beginning of the 21st century, we feel the impact that digital technology and media have on our lives, from personal communication and new forms of entertainment to a radically globalized world.

One of the assets of the Siegen research centre is the participation of projects from a range of disciplines, including media studies, sociology, and informatics. The two projects that joined forces for this volume are a film-history project dealing with the emergence of cinema in Germany around 1900, and an informatics project which develops 21st century methods and tools for all researchers at the centre. Since the use of digital tools is a fairly recent development in media studies, and one which opens up a whole range of new research areas and opportunities for collaboration and result presentation, it might actually be considered to be a kind of media upheaval in its own right. To discuss both potentials and problems arising from the new tools, we invited researchers from all over the world for a workshop. The response to our call for papers, and the lively discussions during the workshop, held in 2007 at Siegen, showed us that we struck a chord when addressing the subject. While many scholars and scientists have already made use of digital tools with considerable success, there still is a great interest in alternative approaches and in the discussion of the scope of possible applications. With this volume, we aim to provide interested parties from both media studies and the informatics an overview of current areas of research and areas of application for digital tools in order to give orientation and to encourage further projects along these lines.

Collecting Data

Perhaps the most basic way to make use of digital tools in media studies is the collecting of data. For decades now, word processing and spreadsheet software have been used to store information electronically for easy access and simple analysis. With database software becoming increasingly flexible and user-friendly, and with huge storage capacity being available at small cost, the collection of large data corpora has become fairly simple, creating an increasing interest in empirical studies. Like in so many other areas, the Internet has played an important role in refining the techniques of data collecting. Firstly, it allows for decentralized data entry: data can be entered from any computer connected to the Internet, and on more than one computer at the same time, allowing for fast and convenient working conditions, especially in a team of researchers. Secondly, the Internet serves as a perfect place for the publication of data collections: it allows world-wide distribution at minimal cost, while at the same time being vastly superior to printed data collections in terms of usability (searching, analyzing etc.) and updating.

Louis Pelletier and Pierre Véronneau describe the genesis of such a database project in some detail, from an original local database created in the mid-1990s to today's Internet-based portal on the history of silent film in Quebec. They show that rapid technical development can even challenge a previous digital project: Is it still necessary to have a "handmade" database-driven index to film-related articles in newspapers and journals of a particular place and time, once these periodicals are available as fully-searchable digital reproductions on the Internet? Pelletier and Véronneau answer in the affirmative, stressing the superiority of human indexation over simple text search routines. One might argue that this is certainly true at the moment, but in the face of current research into digital information retrieval and text mining may look quite differently another decade or so from now. Still, the examples and reasons Pelletier and Véronneau provide, from misprints in the original source to ambiguous names and scores of now obsolete synonyms do not let it seem likely that we will be able to dispense with a kind of human supervision for some time to come. It is a conclusion that will turn up again and again in the course of our discussion.

It must be stressed that the Quebec project greatly profited from the digitalization of its primary sources: it allowed them to expand their corpus and to simplify access to the texts themselves – both for the researchers on the project and the later users of the database. This was made possible by a co-operation between the academic film history project, the local Cinémathèque, and the national library. The example shows that co-operations like these can be extremely fruitful, and they are often seen as a major goal of a networked aca-

demic community. Harry van Vliet, Karel Dibbets, and Henk Gras elaborate on this particular issue a bit further. It is their aim to connect information on cultural events of the past, like film screenings and theatre performances, to the contextual materials still available in archives and museums. Based on two fundamental data collections on Dutch cinema and theatre which in themselves are of great use to historians in the field, the authors are currently working on an ontology to allow database connections between the institutions involved, for what they call "semantic interoperability". Standardization like this is a key issue for any data collecting project that is interested in either incorporating external data or linking the own data with that of others. The problems range from developing rules for indexing personal names or place names to the definition of core databases to which others may refer. As can be seen in the field of libraries, which had to develop standards for cataloguing well before the digital age and were among the first to adapt their rules to the conditions of computers and electronic databases, standardization processes are difficult and can take a lot of time – and may still differ significantly between institutions from different countries or cultural backgrounds.

Even if film and theatre studies can build on standards developed for libraries or other pertinent institutions, the task is still far from elementary, especially where single historical events (e.g., a theatrical performance) are concerned. The Dutch example shows a promising approach by building on an existing collection of core data which is set as a standard and may be referred to by other data collections. While an obvious approach, its success depends greatly on the solution to one substantial problem: It only works if the existing data collection is both sufficiently comprehensive and thoroughly reliable, since it will have to be accepted by all partners of the project. While reliability may be ascertained to some degree by academic supervision, the comprehensiveness of a data collection is naturally limited by the resources available. Especially with international co-operations, existing data collections often turn out to be too limited, most of them having originated from a national context. Examples of non-academic data collections on the Internet show that collections are best accepted that make use of a co-operative data collecting model (e.g., the Internet Movie Database at www.imdb.com, or Wikipedia and its many variants). Thus, projects combining one or more existing core databases with a secure multi-user collaborative platform to form an "Academic Social Web" might be best-suited for academic data collecting in the future. There are plans to advance this concept within the Siegen research centre in the near future, both in theory and practice.

Measuring Success

Collecting data is not the actual aim of scholarship, of course, but rather the means to be able to analyse a particular situation in respect to particular questions. Pelletier and Véronneau describe how the Canadian project started out as a tool for cinema historians to simply find relevant information within the individually unmanageable scores of contemporary periodicals, to become a platform for academic research as well as teaching. Van Vliet, Dibbets and Gras stress the fact that joint databases result in an "enrichment" of information that can be used to re-assess long-held assumptions, e.g. on audience composition in relation to theatrical genres. An area which particularly lends itself to the analysis of collected data is their statistical evaluation. In empirical media studies, the success of a medium, a genre or even a single product is a popular field of research, and it is only natural that this kind of research is made much easier with the help of digital tools, both for collecting the data to be analysed and for the mathematical operations necessary to get at the results.

Today, box-office takings are pretty well documented, so one can get a good idea of the success of particular films or genres compared to others at the same time. Things become more difficult the further one goes back in time. For the first decades in movie history, data is scarce and conclusions will have to be tentative. Even fairly basic data like the number of screenings of a film are difficult to ascertain, and the number of actual viewers is quite impossible to tell. Once the cinema industry gets a little more ordered, and feature films come into the focus of the audience (and of the researcher), things become a little easier. John Sedgwick chose 1930's British cinema to develop his tools for measuring film popularity. His starting point is the rare example of a source which gives insight into audience numbers and revenues at a particular cinema in Portsmouth. To put the data from this source into relation to cinema-going practices in Portsmouth in general, Sedgwick developed a formula to measure success when the available data is limited to newspaper advertisements, resulting in a so-called POPSTAT value for each film. By applying statistical methods to a data collection, Sedgwick discovers discrepancies in audience reception between first-run and other cinemas and offers an approach to explain these discrepancies.

Jaap Boter and Clara Pafort-Overduin look at the same phase of cinema history, in their case in the Netherlands, but focus on the aspect of film distribution rather than reception. By applying Latent Class Analysis to an extensive dataset covering Dutch cinema programmes between 1934 and 1936, they come up with seven clusters of films in relation to the cinemas where they were screened. Contrary to popular belief, however, no clear pattern could be found that relates particular films to particular geographic regions and their

dominant social group – a factor previously thought to be highly formative on Dutch cinema attendance as well as on Dutch culture as a whole. Instead, film distribution seems to follow other rules than the social composition of a cinema's trading area.

The examples of both Sedgwick and Boter/Pafort-Overduin show the potentials of advanced statistical analysis, and at the same time make 'traditional' media scholars aware of the need to get statistical expertise for their empirical research from the outset of their projects. Actually, both papers show the advantages of transdisciplinary research in media studies, in that both John Sedgwick and Jaap Boter are economists who have enhanced the film scholars' understanding of cinema culture with their particular expertise.

Geographical Data

The geographical aspect introduced by Boter and Pafort-Overduin in the context of their socio-geographic analysis is further elaborated by Deb Verhoeven, Kate Bowles, and Colin Arrowsmith. They discuss some of the general questions of using geospatial technologies in cinema studies, drawing on their own project on Greek cinema-going in Melbourne as a case in point. While arguing very much in favour of a geographical approach to audience research, they point out the many challenges that have to be overcome, especially when dealing with historical data. Given the immense amount of data necessary for a thorough analysis of, for example, socio-geographical factors for cinema-going, one of the first questions has to be where to get the data from. Digital data is only available for the most recent times, and even that varies greatly in scope and quality, not to mention compatibility and rights issues. To collect relevant data within a film-related research project will often be far too time-consuming, so that a strategic collaboration with interested parties from other fields of science and scholarship appears to be desirable.

Once these challenges are overcome, researchers are rewarded with new insights into their field that might otherwise have escaped them. This view is reinforced by Michael Ross, Roger Sennert, and Jens Wagner, who present a pragmatic solution to the problem of visualizing the stations of itinerant cinemas in the first decades of cinema. By combining information from a trade journal with a visualization tool based on Google Maps, film researchers were not only able to get a better idea of the actual routes of carnies, and of the areas particularly frequented by itinerant cinemas at the time, but they were also able to enhance their own data by resolving ambiguous identifications, especially with place names referring to more than one place. Considering the possibilities of geospatial tools discussed by Verhoeven et al, the itinerant cinema

tool is only a very early step towards a comprehensive use of its potentials, but as such has proved to be useful already. Again, the lack of available existing data in digital format (including historical maps adjusted to up-to-date carto-graphic standards and historical census data) is the primary challenge for any further development here, and one can only hope that the significance of such data is recognized by funding institutions to allow for their coordinated digi-talisation, for the benefit of geographers, historians and other scholars from all kinds of fields, including media studies.

Old Wine in New Wineskins?

Before we turn to the second part of our collection, which deals with digital tools for film and video analysis and annotation, and the results which can be achieved by this, we would like to address a point raised during our workshop regarding the actual innovation of media research with digital tools. It can be argued, of course, that the kind of research described so far could have been achieved without any digital tools at all: indices can be kept (and sometimes are indeed still kept) on index cards, counting and calculating are possible without computers, and a good printed atlas can provide the information needed to analyse any geographical data. This is certainly true, but it is also somewhat be-side the point. Nobody will argue that the digital tools are doing something completely new and hitherto impossible. However, they can do operations so fast, and can re-do them over and over again with new or updated data, that much research only becomes feasible now that digital tools are available.

As a case in point, another database supervised by the film-history project at Siegen stores information on more than 40,000 films available on the German market between 1895 and 1920. The data ranges from film titles and countries of origin to the films' length and genre, as well as actors and crew members working on a film. The fairly simple database, available on-line at www.fk615.uni-siegen.de/earlycinema/fg/, allows statistical analyses on the film market in Germany for a variety of factors, including the changing domi-nance of some production countries over others within the covered time span, or the development of average film lengths in time and/or in relation to par-ticular genres. With data collections of this size, it instantly becomes obvious that there are great limitations for analysis without the help of digital tools. Considering that the information from the filmography database can be linked to other data, e.g. on programming in particular cinemas, which again might be connected to census data on the social composition of particular areas, the tremendous possibilities of digital tools compared to traditional "handmade" analysis are even more obvious. Digital tools allow for results that in theory

can be achieved by other means, but in practice would almost never have been achieved for pragmatic reasons. At the same time, the new tools require skills for handling them to get the best results from their application.

Tools for Film and Video Analysis

The database-related tools presented so far were more concerned with the context of media and their reception, rather than media products themselves. In the second part of this volume, we turn to tools that help with the analysis of audio-visual media products, i.e. films, videos etc. Classical film scholarship has been very much concerned with "reading" and interpreting films, looking for subtexts, an ulterior meaning or artistic merits. At first, it may seem difficult to imagine how this might be an area for applying digital tools, being so much related to cultural knowledge and human thinking. As we will show, there are plenty of fields where digital tools can at least assist media scholars in their work.

Not surprisingly, the branch of formalist film studies lends itself to digital tools particularly well. In his paper, Yuri Tsivian introduces a self-developed tool which, for all its apparent simplicity, is a powerful tool for anyone who has ever faced the task of counting recurring themes in a film or measuring the lengths of particular items. Tsivian's Cinemetrics software was originally designed, and is most useful for establishing the average shot length of a film, i.e. the average length of film shots between cuts. With Cinemetrics, the actual work of identifying cuts has still to be done manually by the scholar himself, while watching the film, but the simple measuring and counting tool provides all relevant information within seconds after the entering work is completed. What makes Cinemetrics even more important is its design as an Internet-based community tool: results of an analysis can be uploaded into an already impressive database where they immediately become available to other researchers. The collaborative approach allows comparative work on a much larger scale than would be possible for a single researcher.

Ralph Ewerth et al have a more ambitious goal in mind. Their software toolkit Videana was trained to detect film cuts and certain kinds of content in audiovisual material, without human interference. Automatic video analysis has been given much attention lately, and the Videana software scored very impressive results at international contests for shot boundary detections (i.e. cuts) as well as recognition of camera movements, textual elements, and faces. Those sceptic of automatic analyses are able to check Videana's results very easily and correct any mistakes the software may have made. Given recognition rates of 80% to 97% for cut detection, one does not have reason for too

much scepticism, however; especially considering that human annotation reaches very similar scores.

Employing similar technologies from a slightly different angle is the Digital Formalism project introduced by Vera Kropf, Matthias Zeppelzauer, Stefan Hahn, and Dalibor Mitrovic. By focusing on the films by Russian formalist filmmaker Dziga Vertov, the project has an excellent reason to apply tools for detecting formal elements in film to their material. Apart from shot cuts and camera movements, the project aims to reveal patterns in the director's work hitherto unrecognised, using data mining techniques. Furthermore, the researchers expect some help for archives in determining similarities and differences between variant copies of the same film. The technical work is reflected within an advanced film theoretical framework discussing formalism and the digital, and putting the results of the analyses into context.

Another field of application for digitally-assisted formal analysis is introduced by Warren Buckland, who intends to settle a question of authorship by a comparison of directors' film styles. Was *Poltergeist,* Buckland asks, directed by its official director, Tobe Hooper, or its writer and producer, Steven Spielberg? Again, parameters like average shot length, number of camera movements, and shot scales are considered to determine the characteristic style of the two potential directors, which then are matched to the parameters of the film in question. Buckland bases his argument on statistics derived from a relatively small sample: the first 30 minutes from two films of each of the two directors and the first 30 minutes of *Poltergeist*. In view of the good results of automatic detection of at least some of the formal elements examined by Buckland, it is obvious that its use for research of this kind can greatly help to broaden the basis for a statistical analysis and make the results even more convincing.

While the applications presented by Buckland, Kropf et al and Tsivian are primarily academic, Christoph Brachmann et al show a more practical application of film analysis in their paper. The authors have analysed the pattern of trailers for action films and have formalised them in such a way that they are now able to generate similar trailers automatically. The fact that these get comparatively good ratings with test audiences shows the extent to which (a) a formalisation of this particular kind of trailer is possible, and (b) an automatic detection of formal elements can be achieved. Even if film studios will probably stick to their practice of creating trailers manually, the authors' approach is an important step towards processing semantic patterns in audiovisual material.

An important element in the generation of action trailers is the use of sound in films. Leonard Boccia and Peter Ludes elaborate a bit further on this theme, taking into closer view the importance of music in TV broadcasts, here in annual reviews of Brazilian and German TV. The authors introduce the

concept of Key Measures and thereby provide future scientists with helpful instruments to detect particular kinds of sound elements and the semantics involved. At the same time, the intercultural approach by Boccia and Ludes also shows the limitations of any automatic analysis: A precise detection of semantic elements will always depend on the cultural background of the media product analysed, for a particular kind of music might have quite different connotations in different cultural contexts.

Film Annotation

Apart from statistical evaluation, a film analysis – be it automatic or manual – can reveal a wealth of information on a film. The question is how this information is stored to be easily retrieved when needed. The automatic detection tool developed by Ewerth et al already incorporates an annotation tool which allows users to enter free information to any particular frame of the analysed film. For some cases, however, it makes sense to develop more sophisticated solutions.

Margret Schild introduces a solution from the point of view of the librarian and archivist. With the software she describes, it becomes possible to link printed materials like books, advertising materials etc. not only to a film, but to particular scenes of a film. Thus, interested parties can be directed to pertinent literature while watching a film. The software can be used for exhibitions, where the screening of a film can be accompanied by further textual information, or on the Internet, where the viewer of an on-line film, or even of a personally-owned DVD, is pointed to materials relevant to a particular scene.

Rolf Kloepfer's annotation tool AKIRA III brings us back to the question in how far digital tools can assist film scholars in reading and interpreting a film. Kloepfer argues that much of film reception happens on a subconscious level and has to be carefully extracted from the film by a minute examination. His analysis tool allows for manual annotation of films in a way based on the notation of musical scores. When consistently applied, AKIRA III can bring to light patterns of meaning, of repetition, and of coherence which – even though formative on the reception of a film – are so elusive that they often escape the attention of the researcher. Using both a classic and a fairly recent example of film art, Kloepfer successfully shows the potentials of digital tools even for the most sophisticated ways of dealing with media.

Special Thanks

We would like to thank a number of people and institutions involved in the success of this volume. As mentioned earlier on, the volume can be traced back to a workshop held at Siegen in 2007. Annemone Ligensa and Joseph Garncarz provided significant input to the concept of the workshop, and Katrin Barkhausen, Ralph Ewerth, Ingo Köster, Roger Sennert, and Jens Wagner were of particular help in organizing the event. We would also like to thank all the contributors for their lively discussions at the workshop, their co-operation and their patience during the editorial process. Finally, we would like to thank the Deutsche Forschungsgemeinschaft and the University of Siegen for funding and housing the workshop, this publication, and the work of the research centre in general.

Siegen, October 2008

Michael Ross
Manfred Grauer
Bernd Freisleben