New Review of Hypermedia and Multimedia,
Vol. 17, No. 1, April 2011, 73–97

Taylor & Francis
Taylor & Francis Group

A crowdsourcing framework for the production and use of film and television data

GARY GEISLER*, GEOFF WILLARD and CARLOS OVALLE
School of Information, University of Texas at Austin, Austin, TX, USA

(Received 19 July 2010; final version received 30 November 2010)

This paper outlines a framework that would enable the detailed indexing of film and television media through crowdsourcing. By making it easier to generate detailed data about these media on a large scale, fans and scholars can more efficiently produce a wide range of artifacts that reflect their interests in this content. Our development of a test collection included detailed indexing of 12 feature films and 8 television programs. We describe the conditions that make crowdsourcing an ideal approach for accomplishing this work on a larger scale; present a three-level development framework; and discuss how automated indexing, crowdsourcing quality, and copyright concerns might influence continued development of the project. Our framework highlights the potential of both multimedia indexing and crowdsourcing and can serve as a model for others embarking on projects that involve indexing, annotating, or labeling large multimedia collections.

Keywords: Crowdsourcing; Film and television; Indexing; Moving images; Video; Visualization

1. Introduction

The rapid growth of many of today's most popular web sites—Facebook, YouTube, and Wikipedia, for example—can be largely attributed to the work not of the managers or developers of those sites, but to the efforts of what would traditionally be referred to as those sites' *consumers.* That is, much of the actual content the sites provide (articles, videos, photos, messages, comments, and metadata) is contributed by a significant portion of the sites' own customers, acting not only as consumers of the product but producers of it as well. Benkler has said that this "commons-based peer production" (2006, p. 107) model is expanding into "every domain of information and cultural production" (p. 5). Recognizing the shift toward user-led content creation, Bruns (2007) introduced the concept of "produsage" to suggest that traditional notions of content production are becoming outdated and that production and usage increasingly occur simultaneously.

*Corresponding author. Email: geisler@ischool.utexas.edu

New Review of Hypermedia and Multimedia
ISSN 1361-4568 print/ISSN 1740-7842 online
http://www.tandf.co.uk/journals
DOI: 10.1080/13614568.2011.552645

The produsage model highlights the potential value in focusing the expertise, creativity, and energy of the crowd with a system designed to facilitate and leverage its contributions. In recent years, this activity has largely fallen under the term *crowdsourcing* (Howe 2006) to describe how advances in technology are reducing the barriers between amateurs and professionals, and how in an increasingly networked world industries as different as "pharmaceuticals and television" (Howe 2006, p. 179) are able to take advantage of this growing pool of accessible, skilled labor. Group volunteerism on the Web has far deeper roots however, evidenced in particular by Richard Stallman's GNU project[1] (1983) and the ongoing development of the Linux operating system. Linux is an example of what Eric S. Raymond described as a "bazaar" style of software development (1999). Instead of a team overseen by managers working toward a specific release deadline, Linux contributors are free to contribute what they can, when they can. The Linux community is largely self-supporting through its use of online forums, chat rooms, mailing lists, and newsgroups.

French philosopher Pierre Lévy described this process of pooling knowledge to build large-scale projects as collective intelligence. He further theorized a digital communal resource, or "cosmopedia," where global citizens could form groups and share knowledge unfettered from geographic constraints. Henry Jenkins argued

> [O]nline fan communities might well be some of the most full realized versions of Lévy's cosmopedia, expansive self-organizing groups focused around the collective production, debate, and circulation of meanings, interpretations, and fantasies in response to various artefacts of contemporary culture. (2006, p. 137)

A quick look at the discussion page of any contentious Wikipedia topic reveals the arguments, negotiation, and refinement that go into building such a cosmopedia. Indeed, Wikipedia is an exemplary example of Lévy's axiom that "no one knows everything, everyone knows something, all knowledge resides in humanity" (1999 cited Jenkins 2006, p. 139).

As crowdsourcing has developed as a strategy for accomplishing large-scale, collaborative work, several distinct incentive models have emerged. In many crowdsourcing projects, people are paid to perform project tasks; businesses, for example, are using Amazon's Mechanical Turk to distribute routine work assignments such as tagging data, conducting market research, or transcribing audio files to a large pool of people around the world (Hoffmann 2009). Because people are paid for their work, this financial incentive approach often results in quick completion of project tasks.

In situations where it is not feasible or desirable to pay people to contribute work to a collaborative project, approaches that rely on social incentives have proven viable. Luis von Ahn's experiments in crowdsourcing data collection using different types of games as a platform (Games With A Purpose—http://www.gwap.com/gwap/) have shown the feasibility of accomplishing tasks too costly under a financial reward system (2009). In other cases, people participate in a crowdsourced project because of a conscious desire to contribute to the mission of that project. Wikipedia is perhaps the most prominent demonstration of this, growing in less than 10 years to offer 15 million articles in over 200 languages, contributed by "a million registered user accounts and countless anonymous contributors".[2] Many similar examples exist. For instance, although free, high-quality mapping data can be easily accessed and used from Google, Microsoft, and other commercial mapping sources, OpenStreetMap (http://www.openstreetmap.org) has become a very popular crowdsourced initiative, with more than 250,000 people using their local knowledge to contribute edits and improvements (Anderson 2010). Galaxy Zoo was initiated as a crowdsourcing project in 2007 because for some astronomical research existing visual data sets are too large for astronomers alone to examine, and it is often necessary to have multiple independent classifications of the data (Lintott *et al.* 2008); nearly 150,000 people contributed more than 50 million classifications to Galaxy Zoo (http://www.zoo2.galaxyzoo.org) in its first year, work that has been instrumental in several important, unexpected discoveries (e.g. Lintott *et al.* 2008).

The rise of social computing and the attendant trend away from the traditional producer/consumer dichotomy toward a produsage model holds promise for many other types of large-scale, data-intensive projects. The US National Science Foundation, for example, suggested in a 2005 report that new approaches to collecting, indexing, annotating, and sharing data would facilitate advances in the social and behavioral sciences (Berman and Brady 2005). More recently, in their Digital Humanities Start-up grant program[3], the US National Endowment for the Humanities has funded several projects that seek to develop or index data through crowdsourcing. Motivated by these developments, we describe in this paper how crowdsourcing could facilitate the development of a large-scale indexing project in a domain that is of interest to a broad range of people: film and television media. Film and television media is the focus of intense interest by large numbers of fans, scholars, aspiring filmmakers, students, and others. Most recently, fans have relied on wikis to facilitate the generation and sharing of information associated with particular television series or films (i.e. Wookiepedia, Harry Potter Wiki, Wikisimpsons, and numerous others). Of course, there is also a large body of film and television research exploring histories, genres, styles, and themes continually published by the scholarly community in a range of venues. This research, however, has remained largely unchanged in terms of its presentation and dissemination; the vast majority of scholarly material about film and television—an audio-visual medium—continues to take the form of a traditional text-based narrative, occasionally supported by still images extracted from the media being discussed.

As with other domains of interest, technological advances in recent years have created many more possibilities for how film and television can be studied, discussed, taught, and enjoyed. These advances, combined with a public that is increasingly able, willing, and motivated to become not just consumers but also producers, suggest that there is both a need and a means to create a resource that would serve these goals. There are clear signs that "fans" of film and television are now interacting with film and television media not only as consumers but increasingly also as critics and knowledgeable contributors of thoughtful and innovative perspectives on the field. Henry Jenkins claimed fans, as critics, "are so far ahead of either academics or journalists in terms of dealing with television as television," because they are particularly adept are working communally to read and understand character interactions and growth in long-form series that extend over hundreds of hours.[4] Discussion sites centered on film or television media illustrate the degree to which substantial numbers of people are interested in collectively discussing, analyzing, and speculating about specific films or programs. The increased awareness of web-based community efforts brought about by successful attempts at leveraging collective intelligence, such as Wikipedia, has more recently extended to numerous domain-specific projects, arguably lowering the barriers—both actual and perceived—to the production of original, useful perspectives from those who until now have only participated as passive consumers.

One ramification of the growing ease with which communities can form around topics of common interest is the potential for fragmentation (Parameswaran and Whinston 2007). The benefits of collective intelligence are reduced when communities with similar interests are not able to leverage the work each has done. This might occur when communities simply are not aware of each other's existence; increasingly, it is likely to occur when there is no common structure for contributed data and no standard means for sharing work done with that data.

Our primary contribution in this paper is the description of a framework for the development of a large-scale, data-intensive project focused on film and television media. Because large-scale indexing projects involving time-based multimedia are still relatively novel, this framework can serve as a model for those considering similar projects. Our framework emerged from a preliminary project indexing 12 feature films and 8 television programs, including more than 2,500 scenes and 13,000 shots. As described later in the paper, the detailed level of indexing prescribed in our project facilitates a wide range of potential end products, but it is also very labor-intensive. A significant component of our contribution, therefore, is to outline how crowdsourcing could enable projects such as ours to be practically conducted on a scale that would enable them to create a "confluence among communities" (Parameswaran and Whinston 2007, p. 347), and mitigate fragmentation of communities with common interests.

The crowdsourced version of our project consists of three conceptual levels (after Kazman and Chen 2009): (1) a kernel, containing foundational components; (2) a periphery, enabling access to and sharing of data sets, tools, and artifacts (where *artifact*, as used throughout this paper, refers to any user-generated product, such as list of film clips on a given theme, a visualization (Figure 1), or an interactive tool (Figure 2); and (3) a level that enables the "masses" to view and interact with shared artifacts. Before

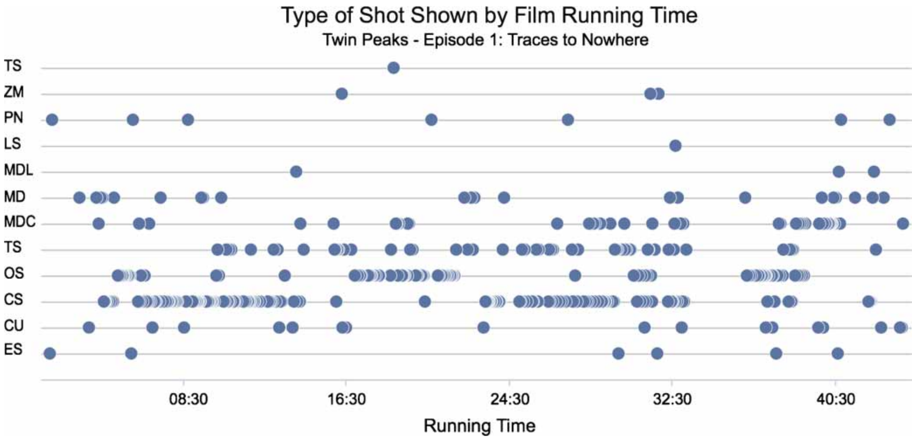

Figure 1. Shot analysis of *Twin Peaks* episode, plotting shot types against running time.

describing the details of our framework, we review related work that helps illustrate the need and feasibility of this project. Following the framework details, we discuss several issues that influence practical implementation of this framework, and conclude by addressing our future work on this project.

Figure 2. Interactive overview of a *Twin Peaks* episode.

2. Related work

2.1 Film and television scholarship

Technologies to access and study film have increasingly granted greater control over media accessibility—from 8mm or 16mm prints in the 1960s and 1970s, to videotapes and laserdiscs in the 1980s, to the DVDs used in the 1990s through today—but they have historically been designed for the consumer market, not a scholarly audience that demands detailed, comparative examination of film. Augst and O'Connor (1999) pointed out the limitations of consumer-oriented technology over a decade ago, noting that the common goals of comparing two sections of film or a section of film with an associated document (e.g. script or journal article) was impractical. The framework we outline in this paper is not targeted specifically at fans looking for a tool to enable rich media analysis, but also at scholars and instructors in film and television studies, communication studies, and associated fields who have for many years intensely studied, analyzed, and written about film and television media. Not only do they have a large amount of domain knowledge to contribute, but also they could utilize the project and its associated artifacts in a teaching environment.

While Augst and O'Connor made the argument for the benefits of the digital environment for studying film, there has since been only limited work in this area. The Virtual Screening Room (VSR), an educational computing project at MIT intended as an "electronic textbook," was an early step in this direction (MIT Center for Educational Computing Initiatives 2006). The VSR was developed as a film browsing and searching system that uses timestamp synchronized transcripts of the film (Ronfard 2004). This system presents a user with an interface for browsing film via short clips, providing analyses oriented toward film editing. MovieBrowser, another project aimed at students of film, uses shot-based segmentation and aggregation techniques to support movie analysis and browsing in a film studies context (Ali and Smeaton 2009). In a study aimed at measuring the benefit of using new technology for this purpose, findings showed improved student outcomes and more positive feedback for the web-based MovieBrowser system compared to using a DVD player for the same film studies tasks (Ali *et al.* 2008). Interestingly, the results showed that students took longer to complete the tasks with the MovieBrowser system; the authors suggest this might be due to students becoming more engaged with their task when using the more granular level of access provided by the MovieBrowser system.

Although not focused on narrative film, a current project at Boston's PBS station, WGBH, serves scholars by providing online access to their in-house generated media content (Michael *et al.* 2009). Their interface design employs visualization techniques made possible by thorough cataloguing. They have found that scholars working with media seek an "intense level" of information about resources (p. 22); WGBH is therefore cataloguing resources at a more detailed level than traditional item-level description, including "sub-item or shot-log level" (p. 22). As described more fully below, this granular level of indexing is also central to our current work.

2.2 Strategies for large-scale multimedia indexing

To support the level of access sought in projects such as the one WGBH is pursuing, it is not sufficient to catalogue resources only at the item level. The metadata for a film that one would find in a library catalogue record, or even at the IMDb, is useful for locating a particular film within a collection, but not for searching within the film itself. The development of a wide variety of new film- or television-based artifacts requires an alternative indexing approach.

Techniques for detailed indexing of multimedia content usually rely on automated content-based detection, which uses low-level (pixel-level) attributes such as color, shape, or texture to infer the existence of higher-level features such as faces, specific types of objects, or settings (Christel and Conescu 2005). Content-based information retrieval has been an active area of research since the early 1990s, but much of the successful work, as demonstrated at the NIST-sponsored TRECVid workshops (Smeaton *et al.* 2004), has involved news, documentary, or sports content. The less consistent structure of fictional film and television media, however, makes automatic indexing techniques less reliable. Likewise, entertainment-oriented content can carry a variety of meanings depending on user interpretation, typically at levels of abstraction that low-level pixel data cannot account for. This *semantic gap* (Smeulders *et al.* 2000) limits the type of successful queries a user can submit to a system that is based solely on automated content-based indexing. At this time, we believe some level of human indexing is necessary to bridge this gap.

Human indexing at the scale ideal for this project entails considerable effort. One way this can be feasible is when an organization has determined the potential benefits are worth the costs of a dedicated team of indexers. Since 1999, The Music Genome Project, which is the database that runs Pandora Radio, has employed a small number of musicians to musically analyze songs based on a number of pre-defined attributes (Westergren 2008). Creating these "musicological fingerprints" takes around 20 minutes for each song, and approximately 10,000–12,000 songs are added to the database each month (Westergren 2008, Pandora 2010). A company called Jinni is creating a similar "genome" for movies by employing "a team of film professionals" to index feature films by characteristics such as mood, tone, and plot elements, aimed at providing a commercial film search and recommendation service for a general audience (Jinni 2010).

These expert-driven sites with a financial reward structure do not fully utilize the contributions of the crowd, however, and the degree to which indexed data is made accessible, editable, and sharable is inevitably tied to the business interests of the company that owns them. More practical models for projects such as ours, where any Internet user can be an indexer, an editor, an artifact contributor, or all of the above, are Wikipedia and similar efforts that rely on social-incentive based crowdsourcing rather than dedicated paid staff.

2.3 Film and television "fandom"

Television and film fans have long used technology to create information sharing hubs, effectively crowdsourcing analysis of their favorite series or movie. Episodic series in particular provide fertile ground for discussion as viewers try to decipher mysteries central to the narrative. *Twin Peaks* (1990), a television series created by David Lynch and Mark Frost, was one of the first shows to exploit Usenet (alt.tv.twinpeaks), and capitalize on networked communication. One fan "provided a detailed sequence of all the narrative events (both those explicitly related and those implied by textual references) and updated it following each new episode", while another "built a library of digitized sounds from the series" (Jenkins 1995, p. 54). With the help of VCRs, fans freeze-framed scenes in their hunt for clues that might reveal the identity of Laura Palmer's killer. The VCR was a tool that enabled viewers to treat the series more like a manuscript, while "the computer net," Jenkins argues, "allowed a scriptural culture to evolve around the circulation and interpretation of that manuscript" (1995, p. 54). Postmodern series like *Twin Peaks* benefitted from this type of detective work by fans as a way of understanding intertextuality created by the writers.

Fan analysis has more recently been accomplished through the help of Wikis, not Usenet. *Lost* (2004), a contemporary show partly inspired by *Twin Peaks*, has an equally devoted fan base operating in a similar manner. The "Lostpedia," a wiki dedicated solely to *Lost* (http://www.lostpedia.wikia.com), is an extensive, fan-curated breakdown of the series, charting character relationships, plot lines, literary techniques, locations, themes, and cultural references. Like *Twin Peaks*, *Lost* has been "widely intertexual," with references to numerous films, television shows, and novels (Lavery 2009, p. 317). As these types of long form narrative series become increasingly complex, they demand more from their audiences. When presented with technologies that allow for the simple sharing of knowledge, fans have demonstrated a willingness to collectively flex their intellectual muscle.

3. A crowdsourcing framework

Although there is a growing body of research on crowdsourcing, thus far it is restricted to studies on specific applications of crowdsourcing, such as to gather labels (Dekel and Shamir 2009) and annotations (Raykar *et al.* 2010), or to evaluate resources (Alonso *et al.* 2008); or studies on particular issues associated with crowdsourcing, such as quality (Mason and Watts 2009), reliability (Raykar *et al.* 2009), and motivation (Huberman *et al.* 2009). The relative novelty of large-scale crowdsourced-based projects, especially those intended to serve at least in part the scholarly community, means that significant empirical research on the obstacles to the successful development of large-scale crowdsourcing projects—and strategies for overcoming those obstacles—has not yet appeared.

There is, however, recent theoretical work that offers potentially valuable strategies that are informed both by open source software development and the more recent emergence of community-based and social networking services

(Kazman and Chen 2009). In this section, we outline our framework for a crowdsourced system for film and television indexing with reference to Kazman and Chen's *Metropolis Model*. In contrast to traditional projects that arise from the bottom up and in which the contributors and requirements are relatively stable, the Metropolis Model speaks specifically to the creation and maintenance of crowdsourced systems that encourage open and collaborative participation. Based on a set of roles and relationships, the Metropolis Model suggests that a crowdsourced application targets three user types, designated by level of involvement: kernel (project developers), periphery (prosumers), and masses (customers). By keeping the levels modular and the kernel tightly controlled, the model allows for unlimited growth from the periphery. Characteristically the periphery and masses levels are open to anyone, producing, contributing, or simply visiting as they see fit, while the kernel is under the purview of a relatively fixed team that develops and manages the underlying system. The system itself is rarely considered "final," and is more often than not in a state of flux based on the emergent behaviors and demands of the users. Likewise, data correctness is a moving target, ultimately resting upon what a community of users deems sufficiently correct.

Because of the demands that follow a commons-based production environment—rapid development, constant change, transparency, and discussion amongst users—our framework follows a set of expectations to best use the wisdom of the crowds in order to index, share, and transform film and television data. The following sections outline that framework, illustrated in Figure 3, using the role levels in the Metropolis Model—kernel, periphery, and masses—as an organizing structure.

3.1 The kernel

Kernel developers create the indexing platform as well as spaces for inter-community engagement, such as forums and mailing lists, to encourage

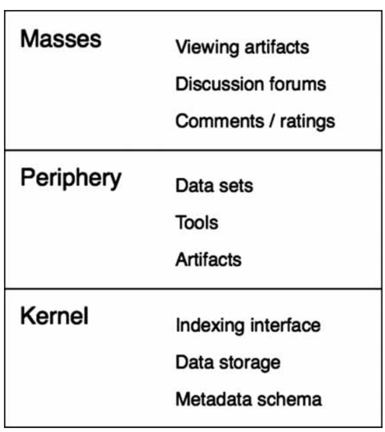

Figure 3. A crowdsourcing framework for the production and use of film and television data.

long-term participation and foster a sense of belonging. Just as the ease with which one can create, edit, and view Wikipedia articles depends on an underlying, mostly invisible foundation (the Media Wiki software and associated components), our kernel consists of several components that provide little direct value to users, but instead enable the development of all system features that do provide value. These components have been developed under the assumption that democratization and acceptance rests heavily on familiarity; our metadata schema is based on conventional film grammar, and our indexing tool has similarities to standalone web-based video players. At the same time, we recognize that in a crowdsourced-based system there will be many useful but unforeseen ideas for modifying and extending the metadata schema and indexing interface; a user forum will provide a means for users in all role levels to offer and discuss new ideas for the kernel components. We fully expect these components to evolve over time as the crowd suggests new ideas and solutions and the kernel developers implement those that are most useful and feasible.

3.1.1 Metadata schema. An increasing number of what might be termed "fan-generated" artifacts involving film and television media have been appearing on the web in recent years. YouTube is a popular ad hoc repository for these new works, including everything from subjective theme analysis to clip compilations aggregating like shots or scenes.[5] While these artifacts demonstrate the interest and willingness to invest significant effort of the non-scholar, they also illustrate how independent efforts of this sort do not scale efficiently. The time and effort one person spends compiling all occurrences of subliminal themes in Alfred Hitchcock's *Psycho* in a YouTube video, for example, cannot be readily leveraged by others who might have interest in the same film and topic; they might have other ideas for analyzing and expressing that interest in another form. In other words, there are no standards or commonly accepted conventions that can be adopted when working with film and television media to produce creative artifacts that would facilitate compatibility, reuse, or integration with work done by others. A fundamental component of our kernel, therefore, is a detailed metadata schema for indexing film and television media that will promote the production, sharing, and reuse of a broad range of artifacts.

There is no established metadata schema that we could simply adopt to serve the goals of this project. Existing schemas intended for cataloguing and describing film and television media (e.g. the Moving Image Collections schema, MPEG-7, and PBCore) served as relevant models in the initial stage of our work, but for different reasons none of these was appropriate for adoption. Instead, we used an iterative process to develop our own project-specific metadata schema (shown in the Appendix), which consists of several distinct high-level components. One component is the project-wide controlled vocabulary elements, defining the vocabularies for elements that are used when cataloguing any film or television program (Table A1). Katz (2005) and Trottier (1998) influenced the choice of these particular terms.

The other metadata schema components are film-specific. One category includes elements that are *not* time-based; that is, they are not synchronized to the media. These include elements such as cast member names, location names, and film-specific controlled vocabularies for other elements.

The remaining metadata elements are time-based metadata elements synchronized to time code in the film and television media. The most fundamental of these are the table of contents elements. These elements fall into a hierarchical structure that follows a common definition of film grammar, consisting of film *sequences*, each of which contains one or more *scenes*, which themselves contain one or more *shots*, as illustrated in Figure 4.

Metadata for the table of contents includes the time code at which each sequence, scene, and shot begins along with additional information that varies depending on the type of structure. Some of this additional metadata (Character and Location) consists of a reference to a previously defined controlled vocabulary element listed in Table A2 (e.g. each character in a scene refers to a Cast » Character element), while other elements, such as Set_type, Time, and Shot_type, consist of a value selected from the project-wide controlled vocabulary shown in Table A1 (e.g. Set_type is one of *Interior*, *Exterior,* or *Montage*).

The remaining time-based metadata elements are sound, song, motif, or script related. Scripts are processed into small chunks, each of which is synchronized to time code in the film and television media and include the speaker of the dialogue and the actual dialogue spoken.

3.1.2 Data storage. For storage of data indexed according to our metadata schema, we have developed both an XML schema and a relational database structure. As mentioned below, XML is the immediate data storage format used by our indexing interface. Because applications can sometimes be easier to develop given database data, we also have implemented a script-based process for transforming source XML files into a relational database structure using the MySQL database management system.

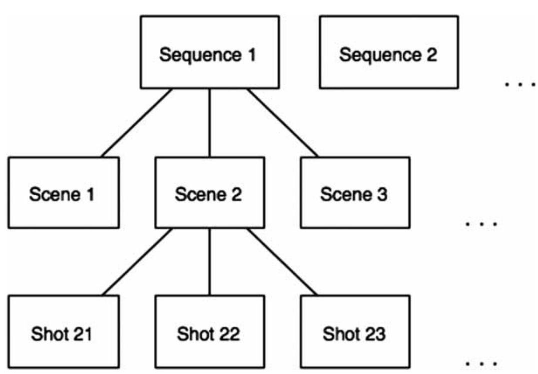

Figure 4. Hierarchical structure used in table of contents.

3.1.3 Indexing interface. Our kernel development has also produced an indexing tool that enables users to index all elements in our metadata schema using an intuitive, web-based interface. Based on a customized version of a product called GLIFOS gmCreator (http://www.glifos.com), the indexing tool (Figure 5) enables, among other features, the synchronization of transcript text and other time-based elements to a multimedia file. This tool is designed for efficient indexing in a collaborative environment; work in-progress can be saved at any time and is stored in an XML file on the server, making it easy for one or multiple users to index a complete film or program over time.

As shown in Figures 5 and 6, much of the indexing is done simply by choosing options from pulldown menus that consist of values from controlled vocabularies.

3.2 The periphery

Supported by the kernel, the Metropolis Model suggests that the next level, the periphery, contributes "the vast majority of end-user value" (Kazman and Chen 2009, p. 81) and is concerned primarily with providing system features to be used by people in the producer/user role. In our framework these are the indexers and artifact producers, which mean periphery features must give people data and tools with which to explore their interests in film and television media. Borrowing from Juno Bacon's idea of a "contributor ramp,"

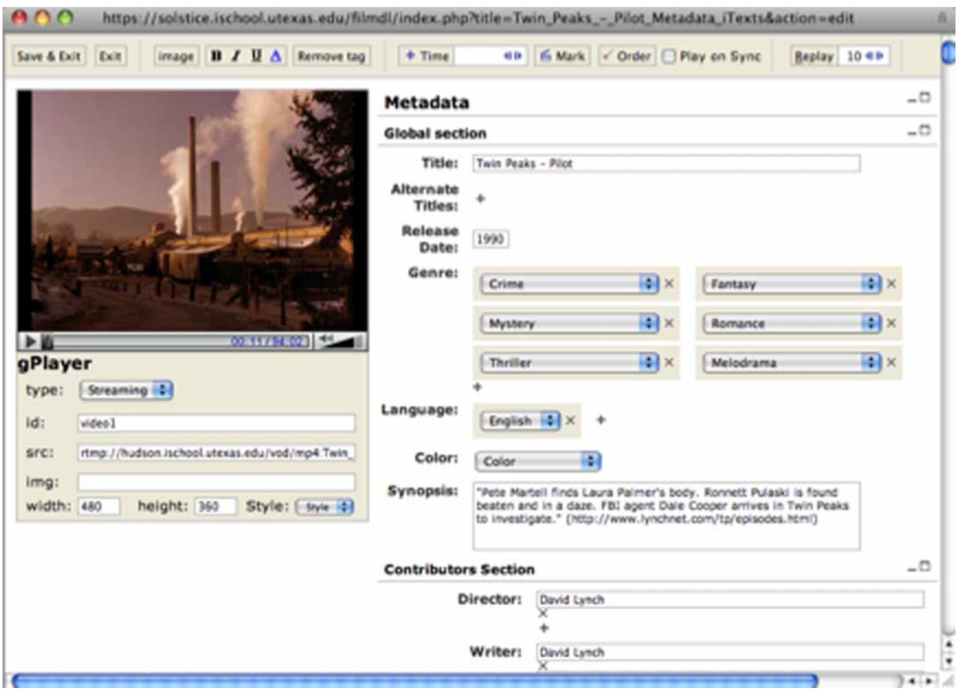

Figure 5. Project interface for indexing global metadata.

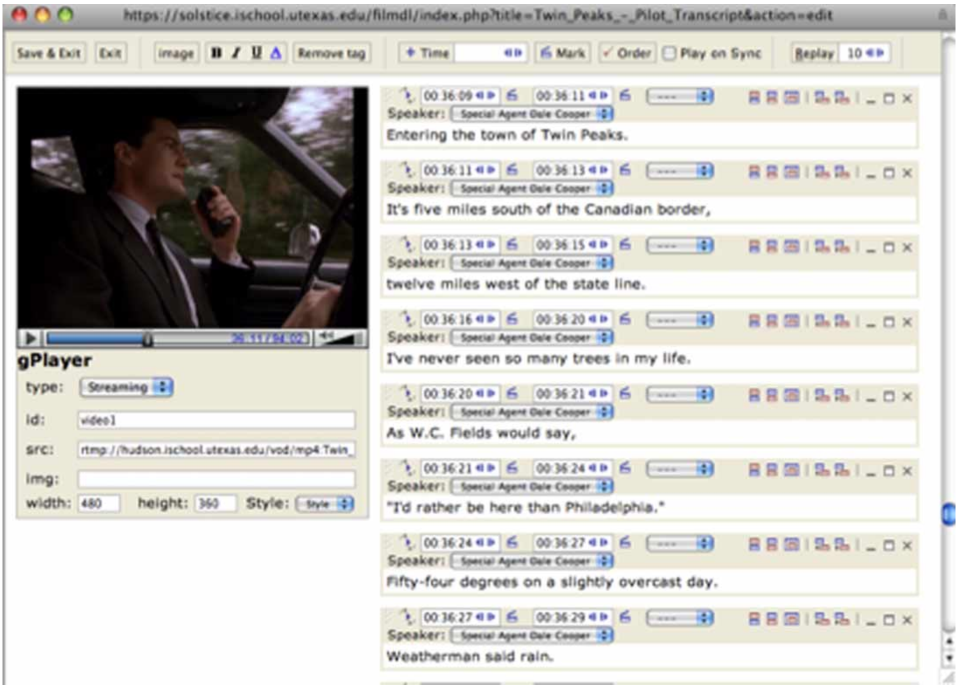

Figure 6. Interface for indexing time-based metadata.

periphery users should be able to identify what they can contribute, where they can contribute, how they can use our tools to do so, and how they can view and share their work (2009, p. 106). This group is not managed in any traditional sense by the kernel, however it is part of a meritocracy.[6] Members who offer frequent contributions, either by indexing media or creating artifacts, will rise in rank and help moderate the system. Together with the built-in communication channels provided by the kernel, we expect this ad hoc peer review system will motivate the periphery.

3.2.1 Data sets. The basis for many of the user-generated artifacts, such as the visualization of characters who co-occur within scenes in the first episode of the American television series Twin Peaks (1990) shown in Figure 7 that we envision being made and shared by "prosumers" via this project will be the granular indexed data enabled by the indexing tool. This data is stored such that it can be accessed at the individual item level (e.g. data set for the film Psycho) or at an aggregate level (e.g. data set for all indexed films directed by Alfred Hitchcock).

Often, there is a specific data format that is more ideal for a particular application or person than another. Our source XML format is a structured, human-readable file format (example snippet below) that enables the indexed data to be relatively easily filtered, transformed, or converted to a range of other data storage formats. After a given film or program has been fully

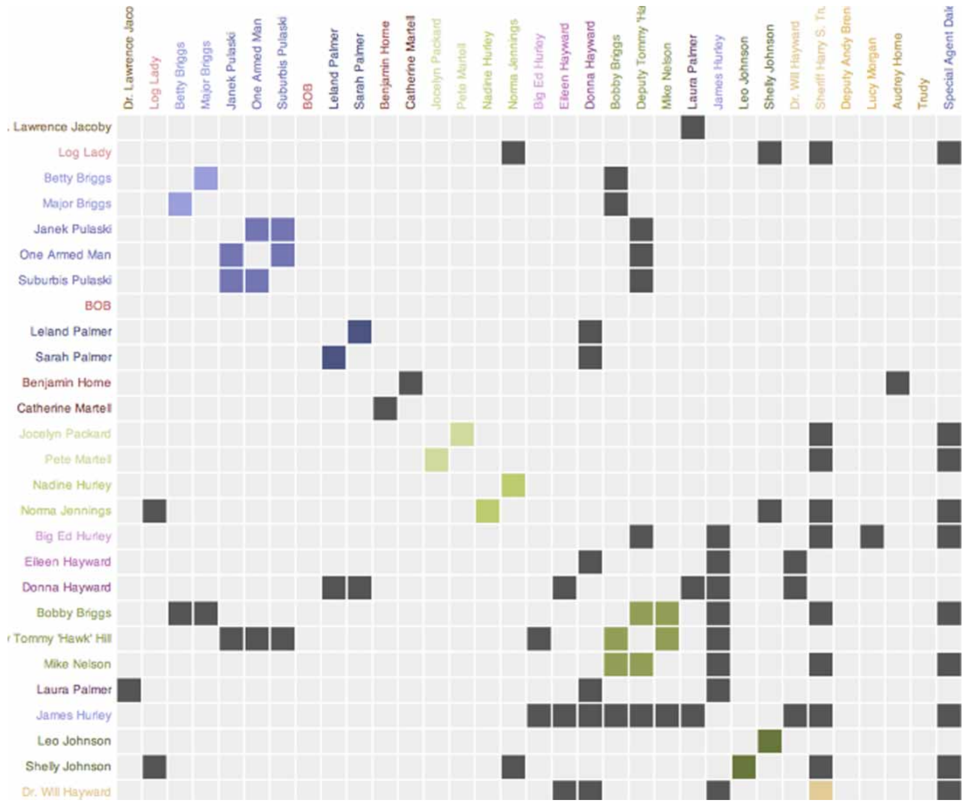

Figure 7. Character co-occurrence interaction matrix, where a filled cell in the matrix shows where characters appear in the same scene.

indexed, users will be able to download its data as a data set in any of XML, comma separated value (CSV), JSON, or MySQL dump file formats.

```
<tm_scene start="00:03:35" set_type="Exterior" time=
"Day" location="Lakeshore By The Blue Pine Lodge">
  <characters>
    <character>Pete Martell</character>
    <character>Laura Palmer</character>
  </characters>
  <shots>
    <tm_shot start="00:03:35" type="Medium Shot"/>
    <tm_shot start="00:03:51" type="Medium Shot"/>
    <tm_shot start="00:03:53" type="Close Shot"/>
 ...
```

3.2.2 Tools. “Tools,” Clay Shirky points out, “don’t get socially interesting until they get technologically boring” (2008, p. 105). While creating an artifact usually entails the use of one or more tools or applications in conjunction with a data set, for most people the tool is just the means to an end. Unfortunately, not everyone with creative ideas for data has the skills to use or develop tools that would help them realize those ideas. Our framework addresses this problem with a provision to make available, and encourage project users to make available, programming scripts and other tools that help transform data sets into artifacts. For example, to use a film's data set to create an arc visualization of character co-occurrence (Figure 8), it is necessary to transform the data into a specific JSON structure that is read by the visualization code. The script we used to perform this transformation can be shared via the site so that others interested in generating this visualization, or variations of it, do not have to reinvent the steps necessary to transform the source data. Over time, contributed scripts and tools can serve as "building blocks" (Geisler 2008) to help users more easily and quickly realize their artifact ideas.

3.2.3 Sharing. With their extensive range of user-contributed data set and visualization artifacts, existing sites such as Many Eyes (http://www.manyeyes.alphaworks.ibm.com/manyeyes/) and visualizing.org demonstrate the value of enabling users to easily upload and share their work for social analysis purposes. While these particular sites emphasize visualization techniques for data exploration, there are myriad possibilities for comparing, summarizing, or viewing a fully indexed film or television data set. Figures 2 and 9 illustrate possibilities for other types of artifacts. In addition to data sets and tools, our framework will enable users to upload and share the artifacts they generate—in fact, it is the artifacts that will likely be the primary interest of users visiting the project, whether in the periphery "prosumer" or the masses "customer" role.

3.3 The masses

Most users in a crowdsourced application fit into the masses role, and can be otherwise identified in our framework as viewers or readers. A Wikipedia survey (Glott *et al.* 2009) found about 63% of their users were considered "readers," and about 24% occasionally contributed. Interacting with and viewing the indexed media content, as well as the user-generated artifacts, indicates the need for transparent access without restriction. Adapting some of the affordances of social media systems, however, can help convert a viewer into a contributor. Discussion forums and wikis lower the barriers to entry for those users already familiar with participatory tools; more importantly they encourage collective analysis. Kazman and Chen assert that "these forums must be made available (typically by members of the kernel) and the periphery encouraged to participate in discussions about the requirements to, in effect, create a community" (2009, p. 9). Because our targeted user base is fairly wide—fans, scholars, aspiring filmmakers, students, and others—the communities we encourage to participate do not fully overlap, and will likely use the site in different ways. Forums are one way these disparate groups can

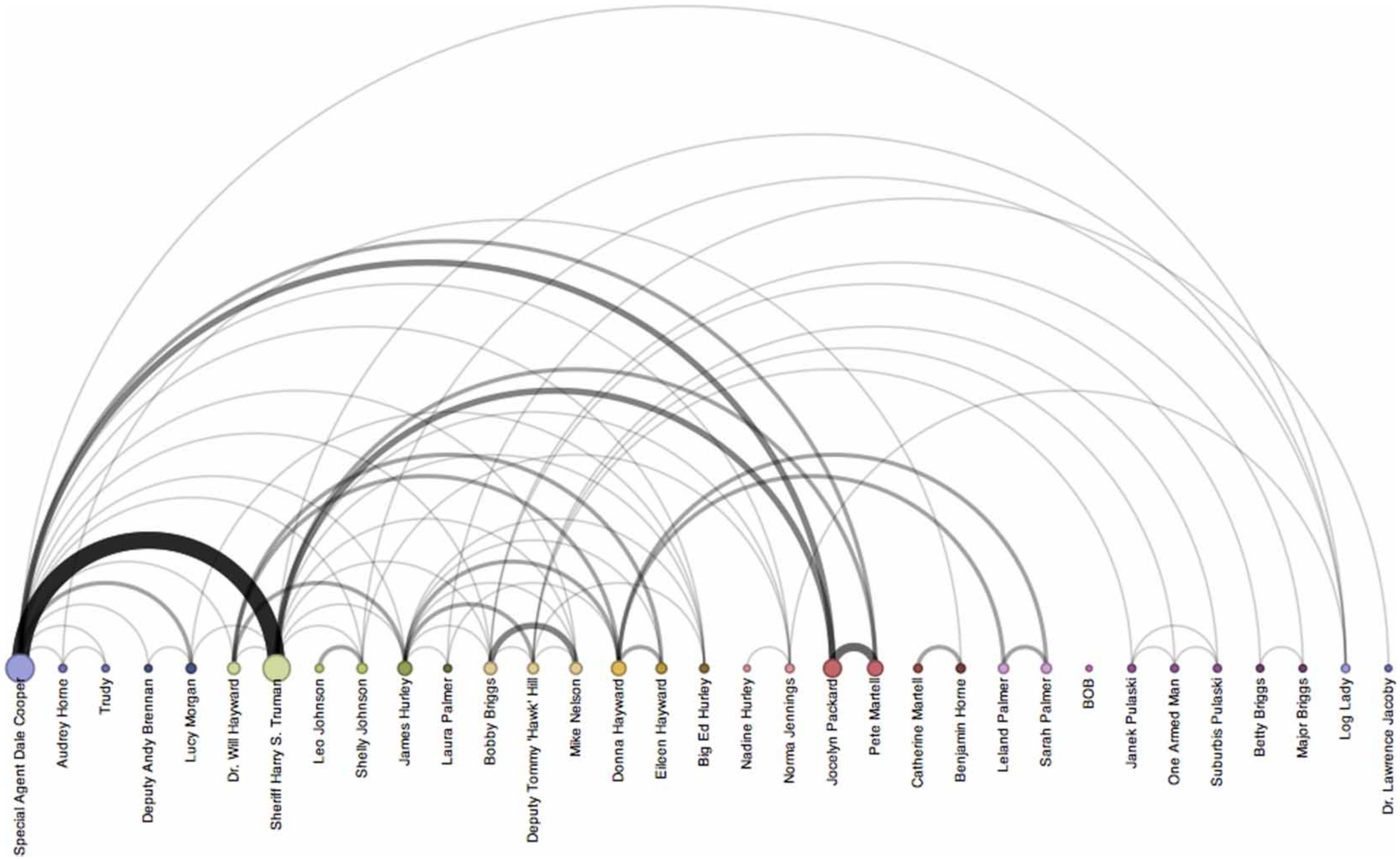

Figure 8. Character co-occurrence arc diagram, where an arc connects characters appearing in the same scene.

be managed. They are also places where ideas can be generated and acted upon, encouraging the passive user to contribute to the site. Enabling comments and ratings of shared artifacts are other possibilities for encouraging participation from the masses.

4. Discussion

The framework outlined above serves as a guide for our implementation of an open, collaborative system for indexing film and television data. Although this implementation is still in a preliminary stage, some components, such as those in the kernel, are functional. While our framework provides a feasible approach to the continued development of a resource we believe would have significant value to a range of audiences, it is an ambitious project, and there are several potential obstacles that could limit its chances for success. In this section, we discuss the most significant of these obstacles, specifically, whether it is realistic to do granular indexing of film and television media on a large scale, whether relying on volunteer indexers will affect the quality of the project data, and whether this work can be done without legal—primarily, copyright—risks.

4.1 Feasibility of large-scale, granular indexing

Human indexing of film and television media at the detailed level we propose is time-consuming. Based on our development of a test collection of 12 feature films and 8 television programs, it can take upward of 10–15 times the

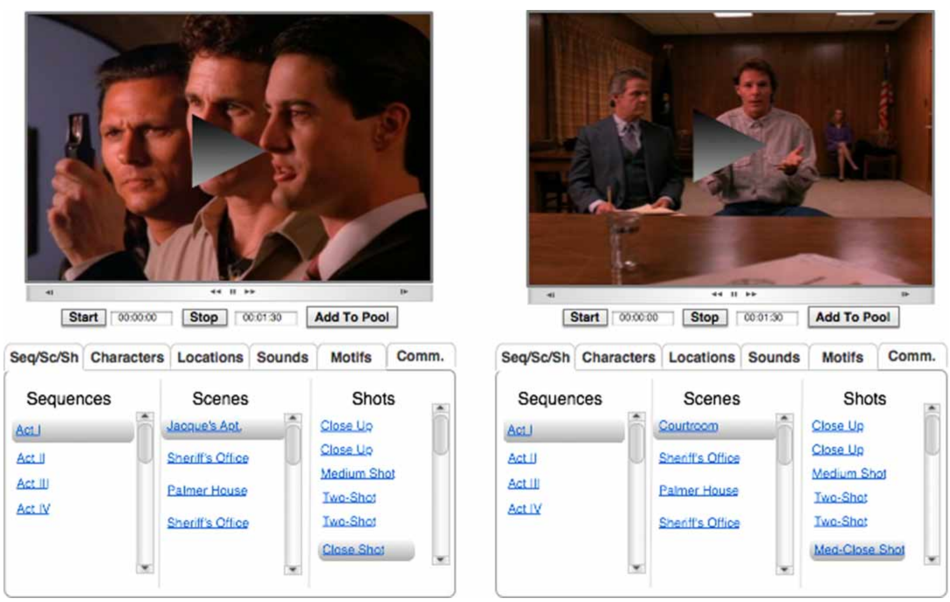

Figure 9. Comparison interface example.

running time of the source material; *Twin Peaks* season one, consisting of a 90-minute pilot and seven 45-minute episodes, took the authors about 85 person hours to index. Although crowdsourcing mitigates this challenge considerably (given enough distribution of effort, no one person has to invest as much time as the authors to benefit from a large corpus of detailed data), it is important to note the potential for some of this effort to be fully or partly automated.

While we are not aware of any current automated technique that can reliably identify the *types* of shots we have outlined in our schema vocabulary, efforts at automated shot boundary detection, based on measuring similarity between adjacent video frames, have been quite successful (Rasheed and Shah 2005) and attempts at automatically classifying scenes as containing specific types of events have more recently shown promise. In Zhai *et al.* (2005), the authors classified film and television clips as *conversation*, *suspense*, and *action*, while Lehane *et al.* (2007) classified portions of content as *dialogue*, *exciting*, or *montage* events. We expect, therefore, that continued advances in automated video processing could provide at least preliminary scene, sequence, and shot divisions for initial "stubs," which could then be corrected by human indexers as necessary. Likewise, years of experiments focused on video-based feature detection as part of the annual TRECVid workshops (Smeaton *et al.* 2008) suggests that the *Set_type* and *Time* index elements in our schema could be automatically populated at a fairly reliable level, and errors would be simple for a human to correct. Longer term, advances in automated techniques could enable the addition of other elements to our model's metadata schema, such as classifying scenes by event type as done by Lehane *et al.* (2007).

Similarly, scores, songs, and sounds have an identifiable fingerprint, and applications like Shazam (http://www.shazam.com/) prove that this fingerprint, for songs at least, is exploitable for music discovery. While these applications show much promise, they are not yet detailed enough to provide practical solutions in this context for identifying scores, songs, and sounds embedded in films or television shows. The improvement of shot, scene, and sequence algorithms over time, along with emergent retrieval techniques for music and sound comparisons, could realistically replace much of the most time-consuming manual work.

A partially automated system would be well supported by crowdsourced editing. Von Ahn's ESP Game might be a useful model for harnessing the crowd to perform a task that computers cannot do. Similar systems (or games) could be developed that present users with short video excerpts based on automatically generated time codes for shots, songs, and sounds. Users would then be tasked with correctly identifying our metadata elements in accordance with our shot vocabulary or supplying tags to song and sound cues.

4.2 Crowdsourcing incentives and quality

Millions of unpredictable contributors working without remuneration bring into question the relationship between financial incentives and data quality.

In a crowdsourced system, users with different skill sets and expertise must cooperate to create high quality work. Perhaps contrary to intuition, this uncontrolled system does not necessarily result in lower-quality results. Wilkinson and Huberman (2007) have found that Wikipedia articles increase in quality as the number of collaborators increase, where "high-quality articles" are associated with a large number of edits and distinct editors. In a study of how financial incentives affect performance of workers using Mechanical Turk, Mason and Watts found that increased payments increased the quantity of work performed but not its quality (2009).

While contributions inevitably will be varying quality, even bad data is better than no data at all, because bad data is a call for improvement. Shirky claims the stubs on Wikipedia are motivators for those who enjoy editing, but reluctantly start new topics: "many more people are willing to make a bad article better than are willing to start a good article from scratch" (2008, p. 122). We believe the film or television equivalent of a stub on Wikipedia would entice users with different skill sets to contribute. Understanding shot types is a different skill from identifying sound design and foley work, but both skill sets can be brought to bear on a system that supports crowdsourced data. Contributors will often lack the expertise of a professional indexer, but recent studies of non-expert annotation (Snow *et al.* 2008, Hsueh *et al.* 2009) and translation quality (Callison-Burch 2009) have shown that within a collaborative system, the work quality of non-experts is not significantly lower than that of the experts, and, as in Wikipedia, it is reasonable to expect that particularly egregious contributions will be quickly identified and corrected.

4.3 Copyright considerations

Providing new forms of access to film and television media has potential legal risks, primarily related to copyright law. This issue is of sufficient complexity that a detailed analysis that covers the entirety of international contexts is beyond the scope of this discussion; however, we will address some considerations and approaches to this problem with particular attention to the US context in which this project is currently being developed.

A key aspect of our framework is the distribution of media indexing to the crowd. As with other ambitious projects requiring detailed work on a large corpus of material, such as Wikipedia or Galaxy Zoo, a social-incentive, crowdsourcing approach is the most feasible alternative to a commercially driven project. The most ideal strategy for enabling large numbers of users to index media content is to provide digital copies of the media on a web server. Users could come to the project site, select a film or television program to index, and use the web-based indexing interface with a directly embedded digital media file. This approach, however, is problematic for several reasons. The creation of a copy and the digital distribution of this film or television media, typically published and in-copyright, could each potentially constitute infringement. Further, although the creation and distribution of a copy for these purposes could be covered by a copyright exemption, if the source material is protected by a technological protection measure (tpm), its circumvention is illegal under US law[7], even for uses that would otherwise be legal. Many commercially released DVDs and digital media are protected by tpms, making their reuse cumbersome. Although there is currently an exemption to this anti-circumvention provision of the law specifically for university professors and media studies and film students[8], it is narrow in scope and is of potentially limited duration (three years, at which point the need for it should again be evaluated by the US Copyright Office if its proponents request its reconsideration).

One can mitigate these risks if the indexing tool directly accesses a DVD or otherwise lawfully acquired local copy of the media. No additional copy would need to be made or distributed, because the user would access his own copy of the work. Although this potentially limits the number of users, since those who wish to use the tool will need their own purchased version of the media, this approach is probably appropriate for a work that is still commercially available.

Another concern involves the metadata generated by users, which could be characterized by a copyright owner as a derivative work of his or her underlying media. Ordinarily, creating derivatives is one of the rights exclusively granted to copyright holders. But *facts* are not ordinarily subject to copyright, and the attempt to claim exclusive rights to create factual metadata in this manner could prove difficult. Furthermore, the character and purposes of these uses support a strong fair use argument, as further discussed below. A related concern involves the possibility that the users who generate the metadata might themselves claim ownership of it, and while facts are not protected by copyright, the selection and arrangement of such facts in a compilation may be protected.[9] Although this risk seems remote, our framework includes the provision that users are asked to agree to allow the use and reuse of their contributions, if any would be copyrightable, so that researchers and contributors can use each other's material for these scholarly purposes.

Under US law, copyright is intended to "promote the progress of science and useful arts"[10] by providing economic incentives for creators. Fair use is the primary method by which educators and researchers can copy and use copyrighted material in fulfilling that purpose. Fair use evaluates four factors as a whole in determining whether or not a given use is infringing. The first factor specifically addresses the purpose of the use, contrasting an unfavorable commercial use with a favorable non-profit educational purpose. Other factors include the nature of the work, the amount of work used, and the effect on the potential market for the work.[11]

Users of this project, intended for educational and research purposes, should be able to make strong fair use arguments. There is a certain amount of risk involved with any venture involving the reuse of copyright works, but the risks should be manageable if reasonable efforts, such as the ones discussed above, are made to protect the legitimate interests of copyright holders.

Additionally, the types of uses themselves strongly support a claim of fair use. In his analysis for the Association of Research Libraries, Jonathan Band suggests that recent appellate decisions support situations in which educators and librarians are likely to assert fair use. More specifically relevant to our project, Band states:

> [E]ducational uses of entertainment products could be viewed as constituting repurposing; as could the display to students of works targeted at scholars. These cases further suggest that educators could buttress their fair use claim by recontextualizing works on course websites through selection and arrangement and the addition of commentary, criticism, annotation, and student reactions. (2007, p. 17)

5. Conclusion and future work

Increasing awareness of and interest in participatory culture, coupled with advances in technology that make it easier to foster collaborative efforts, is resulting in a growing range of projects whose users are also its producers. We have outlined a framework for developing a collaborative project intended for audiences interested in film and television media. This project relies on crowdsourcing to produce detailed media indexing, which can then be used to generate, share, and discuss a broad range of artifacts. We have developed a significant portion of the core, or kernel, function of the project. Work still to be done includes building out the features that will enable others to fully participate in the project including a public web site.

As we further develop this project, an important thread of our work is obtaining feedback from our intended audience. This includes formal user research within the academic community to find out more about their use of existing resources when doing research and teaching with film and television media, and in particular how this project might be developed to encourage participation from the scholarly community. At the same time, we are equally interested in obtaining reactions and suggestions from the broader community via the project web site[12] as it evolves.

We are also still exploring the range of artifacts that can be produced from indexed data. There is a significant amount of previous research on developing tools and systems for working with collections of digital video, but much of this previous work is based on the availability of a more limited range of time-based metadata. The possibilities that exist from the depth of indexing we have produced in our current work leave much to explore.

Acknowledgements

This work was funded in part by a John P. Commons Teaching Fellowship from the University of Texas at Austin School of Information. We also thank Niki Arroyave, Rodrigo Arias, and the team at Glifos for their work on the indexing software used in this project. Finally, we acknowledge the valuable effort and feedback of all the UT iSchool students who have contributed to this project.

Notes

[1] http://www.gnu.org/gnu/gnu.html
[2] http://www.en.wikipedia.org/wiki/History_of_Wikipedia
[3] http://www.neh.gov/grants/guidelines/digitalhumanitiesstartup.html
[4] http://www.boingboing.net/2010/04/16/reducing-the-worlds.html
[5] http://www.youtube.com/user/robag88; http://www.youtube.com/watch?v=aZZxRhJTG4k
[6] Apache, Wikipedia, Slashdot, and Discogs, among many others, are built as meritocracies with different variations on how users are rewarded for their expertise.
[7] 17 USC Sect. 1201
[8] 37 CFR 201.40 (b)(1)
[9] Feist Publications, Inc. v. Rural Telephone Service Company, Inc., 499 US 340 (1991).
[10] US Const. art. I, Sect. 8, cl. 8
[11] 17 USC Sect. 107
[12] http://www.filmanalysis.org

References

N.M. Ali and A.F. Smeaton, "Are visual informatics actually useful in practice: A study in a film studies context", in *First International Visual Informatics Conference, IVIC 2009*, 11–13 November 2009, Kuala Lumpur, Malaysia, New York: Springer, pp. 811–821, 2009.

N.M. Ali, A.F. Smeaton, H. Lee and P. Brereton, "Developing a MovieBrowser for supporting analysis and browsing of movie content", in *iHCI 2008—Irish Human Computer Interaction Conference*, 19–20 September 2008, University College Cork, Cork, Ireland, pp. 1–6, 2008.

O. Alonso, D.E. Rose and B. Stewart, "Crowdsourcing for relevance evaluation", *SIGIR Forum*, 42(2), pp. 9–15, 2008.

N. Anderson, 2010. OpenStreetMap: Crowd-sourcing the world, a street at a time. Ars Technica. Available online at: http://www.arstechnica.com/open-source/news/2010/06/crowd-sourced-world-map.ars (accessed 12 July 2010).

B. Augst and B.C. O'Connor, "No longer a shot in the dark: Engineering a robust environment for film study", *Computers and the Humanities*, 33(4), pp. 345–363, 1999.

J. Bacon, *The Art of Community: Building the New Age of Participation*, Sebastopol, CA: O'Reilly Media, Inc., 2009.

J. Band, 2007. Educational fair use today. Association of Research Libraries. Available online at: http://www.arl.org/bm ~doc/educationalfairusetoday.pdf (accessed 10 July 2010).

Y. Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, New Haven, CT: Yale University Press, 2006.

F. Berman and H. Brady, 2005. Final report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences. San Diego Supercomputer Center. Available online at: http://www.ucdata.berkeley.edu/pubs/CyberInfrastructure_FINAL.pdf (accessed 28 October 2010).

A. Bruns, "Produsage", in *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition*, 13–15 June 2007, New York, New York: ACM, 2007.

C. Callison-Burch, "Fast, cheap, and creative: Evaluating translation quality using Amazon's Mechanical Turk", in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 6–7 August 2009, Singapore, Singapore: World Scientific, pp. 286–295, 2009.

M.G. Christel and R.M. Conescu, "Addressing the challenge of visual information access from digital image and video libraries", in *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries*, 7–11 June 2005, Denver, CO, New York: ACM, pp. 69–78, 2005.

O. Dekel and O. Shamir, "Vox populi: Collecting high-quality labels from a crowd", in *Proceedings of the 22nd Annual Conference on Learning Theory,* 18–21 June 2009, Montreal, Quebec, 2009.

G. Geisler, "Building blocks for rapid development of information seeking support systems". Position paper for National Science Foundation sponsored *Information Seeking Support Systems Workshop*, 26–27 June 2008, Chapel Hill, NC, 2008.

R. Glott, P. Schmidt and R. Ghosh, 2009. Wikipedia survey—first results. Wikipedia. Available online at: http://www.upload.wikimedia.org/wikipedia/foundation/a/a7/Wikipedia_General_Survey-Overview_0.3.9.pdf (accessed 15 July 2010).

L. Hoffmann, "Crowd control", *Communications of the ACM*, 52(3), pp. 16–17, 2009.

J. Howe, 2006. The rise of crowdsourcing. *Wired,* 14(6). Available online at: http://www.wired.com/wired/archive/14.06/crowds.html (accessed 9 July 2010).

P.Y. Hsueh, P. Melville and V. Sindhwani, "Data quality from crowdsourcing: A study of annotation selection criteria", in *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*, 31 May, Stroudsburg, PA: Association for Computational Linguistics, 2009.

B.A. Huberman, D.M. Romero and F. Wu, "Crowdsourcing, attention and productivity", *Journal of Information Science*, 35(6), pp. 758–765, 2009.

H. Jenkins, "Do you enjoy making the rest of us feel stupid? alt.tv.twinpeaks, the trickster author, and viewer mastery", in *Full of Secrets: Critical Approaches to Twin Peaks*, D. Lavery, Ed., Detroit, MI: Wayne State University Press, pp. 51–69, 1995.

H. Jenkins, *Fans, Bloggers, and Gamers: Exploring Participatory Culture*, New York: New York University Press, 2006.

Jinni, 2010. The movie genome. Available online at: http://www.jinni.com/movie-genome.html (accessed 25 May 2010).

E. Katz, *The Film encyclopedia: The Most Comprehensive Encyclopedia of World Cinema in a Single Volume*, (5th ed)., New York: Collins, 2005.

R. Kazman and H.-M. Chen, "The metropolis model: A new logic for development of crowdsourced systems", *Communications of the ACM*, 52(7), pp. 76–84, 2009.

D. Lavery, "*Lost* and long-term television narrative", in *Third Person: Authoring and Exploring Vast Narratives*, P. Harrigan and N. Wardrip-Fruin (Eds), Cambridge, MA: The MIT Press, pp. 313–322, 2009.

B. Lehane, N.E. O'Connor, H. Lee and A.F. Smeaton, "Indexing of fictional video content for event detection and summarisation", *EURASIP Journal on Image and Video Processing,* pp. 1–15, 2007.

C.J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M.J. Raddick, R.C. Nichol, A. Szalay, D. Andreescu, P. Murray and J. van den Berg, "Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey", *Monthly Notices of the Royal Astronomical Society*, 389(3), pp. 1179–1189, 2008.

Lost, 2004. Television Program, ABC Studios.

W. Mason and D.J. Watts, "Financial incentives and the 'performance of crowds'", in *HCOMP '09: Proceedings of the ACM SIGKDD Workshop on Human Computation*, 28 June 2009, Paris, New York: ACM, pp. 77–85, 2009.

C. Michael, M. Todorovic and C. Beer, "Visualizing television archives", *Bulletin of the American Society for Information Science and Technology*, 35(5), pp. 19–23, 2009.

MIT Center for Educational Computing Initiatives, 2006. The virtual screening room. Available online at: http://www.ceci.mit.edu/projects/virtual_screening_room/index.htm (accessed 20 May 2010).

Pandora, 2010. Pandora FAQ. Available online at: http://www.blog.pandora.com/faq/ (accessed 25 May 2010).

M. Parameswaran and A.B. Whinston, "Research issues in social computing", *Journal of the Association for Information Systems*, 8(6), pp. 336–350, 2007.

Z. Rasheed and M. Shah, "Detection and representation of scenes in videos", *IEEE Transactions on Multimedia*, 7(6), pp. 1097–1105, 2005.

V.C. Raykar, S. Yu, L. Zhao, A. Jerebko, C. Florin, G. Hermosillo Valadez, L. Bogoni and L. Moy, "Supervised learning from multiple experts: Whom to trust when everyone lies a bit", in *Proceedings of the 26th Annual International Conference on Machine Learning*, 14–18 June, New York: ACM, pp. 889–896, 2009.

V.C. Raykar, S. Yu, L. Zhao, G. Hermosillo Valadez, C. Florin, L. Bogoni and L. Moy, "Learning from crowds", *Journal of Machine Learning Research*, 11, pp. 1297–1332, 2010.

E.S. Raymond, *The Cathedral & the Bazaar*, Sebastopol, CA: O'Reilly, 1999.

R. Ronfard, "Reading movies: An integrated DVD player for browsing movies and their scripts", in *Proceedings of the 12th Annual ACM International Conference on Multimedia*, 10–16 October, New York, New York: ACM, pp. 740–741, 2004.

C. Shirky, *Here Comes Everybody: The Power of Organizing Without Organizations*, New York: Penguin Press, 2008.

A.F. Smeaton, P. Over and W. Kraaij, “TRECVID: Evaluating the effectiveness of information retrieval tasks on digital video”, in *Proceedings of the 12th Annual ACM International Conference on Multimedia*, 10–16 October, New York, New York: ACM, pp. 652–655, 2004.

A.F. Smeaton, P. Wilkins, M. Worring, O. de Rooij, T.S. Chua and H. Luan, “Content-based video retrieval: Three example systems from TRECVid”, *International Journal of Imaging Systems and Technology*, 18(2–3), pp. 195–201, 2008.

A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, “Content-based image retrieval at the end of the early years”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), pp. 1349–1380, 2000.

R. Snow, B. O’Connor, D. Jurafsky and A.Y. Ng, “Cheap and fast—but is it good?: Evaluating non-expert annotations for natural language tasks”, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing,* 25–27 October, Stroudsburg, PA: Association for Computational Linguistics, pp. 254–263, 2008.

D. Trottier, *The Screenwriter’s Bible: A Complete Guide to Writing, Formatting, and Selling Your Script*, (3rd ed)., Los Angeles: Silman-James Press, 1998.

L. von Ahn, “Human computation”, in *DAC ’09: Proceedings of the 46th Annual Design Automation Conference*, 26–31 July, San Francisco, New York: ACM, pp. 418–419, 2009.

T. Westergren, 2008. The story of Pandora so far from founder Tim Westergren. Available online at: http://www.intruders.tv/en-tech/the-story-of-pandora-so-far-from-founder-tim-westergren/ (accessed 25 May 2010).

D. Wilkinson and B. Huberman, “Cooperation and quality in Wikipedia”, in *Proceedings of the 2007 International Symposium on Wikis*, 21–23 October, Montréal, New York: ACM, pp. 157–164, 2007.

Y. Zhai, Z. Rasheed and M. Shah, “Semantic classification of movie scenes using finite state machines”, *IEE Proceedings of Vision, Image and Signal Processing*, 152(6), pp. 896–901, 2005.

Appendix 1

The tables below illustrate our framework’s initial metadata schema. Elements marked with an asterisk in these tables are repeatable fields (i.e. can have multiple values).

Table A1. Project-wide controlled vocabulary elements.

Element	Vocabulary Values
Set_type	Interior, Exterior, Montage
Time	Day, Night, Dusk, Dawn, Unknown
Shot_type	Close Up, Close Shot, Medium Close Shot, Medium Shot, Medium Long Shot, Over-the-Shoulder Shot, Point-of-View, Two-Shot, Long Shot, Establishing Shot, Tracking Shot, Insert, Pan, Zoom, Title Card

Table A2. Non-time-based metadata elements.

Global elements	Contributors	Cast	Locations
Title	Director*	Character Name*	Location*: Name, Description
Alternate Title*	Writer*	Actor: Name, Sex, Headshot image	
Release Date	Producer*		
Genre*	Production Company*		
Language*	Cinematographer*		
Color	Film Editor*		
Synopsis	Costume Designer*		
	Composer*		
	Sound Designer*		
	Production Designer*		
Songs	Sounds	Motifs	
Song*: Title, Author, Performer	Sound*: Title, Type, Description	Motif*: Title, Type, Description	

Table A3. Table of contents metadata elements.

Sequence	Start_time, Title
Scene	Start_time, Set_type, Time, Location Character*
Shot	Start_time, Shot_type

Table A4. User-defined metadata elements.

Sound	Start_time, Stop_time, Sound, Note
Motif	Start_time, Stop_time, Motif, Note
Song	Start_time, Stop_time, Song
Commentary	Start_time, Stop_time, Title, Description Reference*: Start_time, Stop_time