Dan Albertson and Amanda B. Nickerson
University at Buffalo, The State University of New York, Buffalo, NY, USA

DIGITAL SOCIAL SERVICES: FROM DATA AGGREGATION TO CULTURALLY COMPETENT CONTENT (Paper)

Abstract: A discussion is provided where various data-intensive efforts of a funded digital social services project are reported on. The goal of the project is to develop a digital library and built-in peer-to-peer support features that serve highly diverse audiences and users of New York State. Current work has already provided a number of insights regarding data aggregation and processing needed for making progress toward culturally competent content for digital social services. Those insights are detailed here, along with planned user-centered evaluations and future data-driven and theoretical research streams. The work described here can raise awareness regarding data requirements for digital social services.

Résumé:

1. Introduction

Digital social services employ computing technologies to provide public outreach and/or support services to communities. These projects are many times funded by local, state, and/or federal government agencies in order to help address a particular problem, discrepancy, or need. In order to serve the general public, digital social services must be able to reach, communicate, and be used effectively by diverse populations.

The needs and requirements for cultural competence (in addition to accessibility and usability) are becoming more and more recognized (Ilomäki et al., 2016). Culturally competent information is that which can be interpreted effectively and avoid misuse or misinterpretation across different cultures, languages, and/or backgrounds of individuals. Cultural competence is imperative for many different types of information services (e.g. governmental, health care, education) including those that are more traditional and increasingly in digital contexts where considerations for data processing and management are needed (Chang et al., 2004).

Cultural competence within digital social services is the topic as presented; a project is used as a case to describe different types of data tasks, content needs, and roles of information professionals. While the present paper reflects on practical aspects of data work, future data-driven and theoretical research streams will also emerge from the project, as also described here.

2. Data Tasks and Content Needs

Efforts of an ongoing funded project have resulted in specific types of data tasks and requirements for progressing toward culturally competent content, a necessity of the project.

Project Overview

The project entitled *Multimedia Peer-to-Peer Abuse Prevention* has been funded by the New York State Developmental Disabilities Planning Council (DDPC) in order to create a digital library and online peer-to-peer support feature that will provide educational information, resources, and other digital support features to individuals with intellectual and developmental disabilities who are experiencing various forms of exploitation and abuse. High quality information resources are being identified that will help individuals build personal capacity. The resources and tools developed throughout this project will be accessible on both standard computers and mobile devices, and will be fully accessible to individuals with different abilities and from diverse cultural and linguistic backgrounds. Digital social services with online support mechanisms can provide a number of different benefits to users including a sense of empowerment, reduced loneliness, improved knowledge, decision-making skills, social well-being, and optimism (Barak et al., 2008; van Uden-Kraan et al., 2009). However, a number of tasks and considerations are necessary for supporting culturally competency for the digital content.

Content Aggregation and Data Processing

Many times, the emphasis of digital social services is the aggregation of quality content from multiple sources (as opposed to content creation). The primary benefit of aggregating content is that it is difficult for people who need assistance from social services to find the information they need, as resources are often siloed, the authority and significance of the information is unclear, the information is not objective (but biased), resources may be out of date and/or incomplete, and/or users may not feel confident in their abilities to search and find information online (Albertson & Ju, 2016; Flanagin & Metzger, 2007; Metzger, 2007).

A primary objective for content aggregation is to collect diverse types and formats of information from high-quality sources, with particular emphasis on multimedia so that users with different abilities and literacies can use and understand the content. Here, a vetting system needs to be in place, particularly considering its use by potentially vulnerable audiences. Ultimately, for digital social services, all resources and content provided – regardless of information type – should strive to demonstrate cultural competence, which requires a number of steps.

First, data from aggregated content must be extracted and processed. Textual data can be extracted from source files of textual documents and from time-based media, i.e. videos, using closed captioning or automatic speech recognition (ASR) outputs. Once full text has been extracted, additional layers of processing can be applied to initiate progress toward cultural competency, such as parsing textual documents and formatting transcripts in structures compatible with streaming video and players.

Next, translation of all data – regardless of format – will be a major step in data processing. To prioritize which languages, many times those most commonly spoken in particular areas have been identified through state or local surveys and published in publically accessible data. Such public data can serve as an indicator for which languages to prioritize the content of digital social services projects.

Most importantly, decisions will have to be made on the translation approach based on the number of different languages and the volume of content. Further, there is an obvious tension between automated translation approaches versus manual, particularly when a sizable collection is involved. Automated translation approaches are available through use of programming APIs, which offer speed and efficiency for processing large volumes of data into many different languages, with modest accuracy in syntax. On the other hand, manual translation (particularly for certified content) offers semantic and syntax accuracy. To summarize, automated translation runs a much greater risk of miscommunicating or misinterpreting the semantics of messages, while manual or certified translation of even a modest dataset in several languages will result in a substantial cost. Therefore, while ideally all translated content provided by digital social services would be certified, a balance between what can be systematically translated versus that which cannot needs to be found.

There are other factors which are significant for data aggregation and processing. These include ensuring data extraction and parsing produce low-level content, or that with basic language structures, information that is concrete or factual, with limited abstraction and/or required interpretation on behalf of the user, preservable document structures which are compatible with different writing systems, and other important aspects of digital literacy. Effort given during the vetting process (when aggregating resources) will in turn aid in these qualities. To summarize, the primary steps to assist with cultural competency in digital social services include:

- Aggregating multimedia (as a primary resource) and content vetting
- Data extraction from source files and content
- Translation and other data processing
- Testing content in user interfaces and structured documents

Once these steps of data aggregation and processing have been conducted, whether for textual or multimedia, appropriate techniques for organizing, indexing, or retrieving information can be tested and applied to enable access and classification of resources; these aspects of the project will be reported on at a later date.

3. Opportunities for Information Professionals and Education Programs

Information professionals in particular are prepared to contribute to and work for such public and social services in a number of ways. In addition to data and user-centered design skills, information professionals are many times provided with experience in interfacing with the public as part of their education and/or on the job training. Thus, information professionals directly experience the need and importance of cultural competency on a regular basis. Data and public service skills coalesce nicely into an applicable skill set for digital social services. As a result, there are opportunities to include courses within degree programs that examine both digital data (processing, management, and curation) skills and cultural competency – together – which would be a unique contribution of an information studies degree (Ilomäki et al., 2016; Chang et al., 2004).

4. Future Work and Research Streams

Beyond development and data management, future work will also include both evaluation and research. Evaluations will be conducted to measure progress toward project-level goals and

expected outcomes. Different research streams will also emerge which will enable researchers to examine both data-focused and theoretical research topics.

Evaluation

The project will first include formative evaluation in terms of how the digital library and peer-to-peer support feature should be developed and designed to best support users and their needs. This level of formative evaluation will include focus groups with users from the target audience who will be recruited and asked to provide input on designs, online tools, and resources that they find useful. Initial focus groups are particularly important for the project considering the target audience and sensitivity of the topics to be covered. Therefore, user input is critical from the inception of the project.

Further activities will include evaluations to measure the effectiveness of the digital library, particularly as it adheres to cultural competency. The planned evaluations will be useful for future digital library projects, which traditionally emphasize usability and other quantitative assessments. While usability is critical (and will be part of the ongoing evaluation measures of the current project) the work as presented here demonstrates the need to measure other issues and requirements of digital social services as they relate to cultural competence (e.g trustworthiness, confidence, confidentiality, and sensitivity). This work provides a new perspective of digital library evaluation in showing how digital projects can benefit from qualitative evaluations of use.

Research

Different research streams can emerge from the project. First, the project can serve as a basis to examine and conduct research that focuses on data and digital content. Content curation and culturally competent information, as examined together, can comprise a notable research area. Here, researchers can examine how informational topics can be organized and guided to ultimately support and improve learning of users from highly diverse cultural backgrounds and abilities. Additionally, data ethics issues can be examined including that of confidentiality, security, moderation of user-contributed information, user buy-in, user engagement on sensitive topics, and others.

Secondly, theoretical research can also emerge where the digital library and community can form a case for examining different theories that pertain to online communities, social learning and behaviors, and other phenomena. A variety of theories are applicable and relevant to case studies and/or a virtual ethnographies, which can be used a lens to analyze data or adapt new theory for online societies.

5. Acknowledgement

Reference List:

Albertson, D. & Ju, B. (2016). Perceived self-efficacy and interactive video retrieval. *Journal of Documentation, 72*(5).

Barak, A., Boniel-Nissim, M., & Suler, J. (2008). Fostering empowerment in online support groups. *Computers in Human Behavior, 24,* 1867-1883.

Chang, B. L., Bakken, S., Brown, S. S., Houston, T. K., Kreps, G. L., Kukafka, R., ... & Stavri, P. Z. (2004). Bridging the digital divide: reaching vulnerable populations. *Journal of the American Medical Informatics Association, 11*(6), 448-457.

Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of Web-based information. *New Media & Society, 9*, 319–342. doi:10.1177/146144480707501

Ilomäki, L., Paavola, S., Lakkala, M., & Kantosalo, A. (2016). Digital competence–an emergent boundary concept for policy and educational research. *Education and Information Technologies, 21*(3), 655-679.

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the Association for Information Science and Technology, 58* (13), 2078-2091. doi:10.1002/asi.20672

van Uden-Kraan, C. F., Crossaert, C. H. C., Taal, E., Seydel, E. R., & van de Laar, M. A. F. J. (2009). Participation in online patient support groups endorses patients' empowerment. *Patient Education and Counseling, 74*, 61-69.