Creating Public Value by Democratizing the Ecosystem of Human Service Providers

Charalampos Chelmis
University at Albany, SUNY
Department of Computer Science
cchelmis@albany.edu

Mengfan Yao
University at Albany, SUNY
Department of Computer Science
myao@albany.edu

ABSTRACT

Human service providers play a critical role in improving well–being in the United States. However, little is know about (i) how service seekers find the services they are looking for by navigating among available service providers, and (ii) how such organizations collaborate to meet human needs. In this paper, we report the first outcomes of our ongoing project. Specifically, we first describe a data acquisition engine, designed around the particular challenges of capturing, maintaining, and updating data pertaining to human service organizations from semistructured Web sources. We then proceed to illustrate the potential of the resulting comprehensive repository of human service providers through a case study showcasing a mobile app prototype designed to provide a one–stop shop for human service seekers.

CCS CONCEPTS

• **Information systems** → **Web crawling**; *Information systems applications*; • **Human-centered computing** → *Graphical user interfaces.*

KEYWORDS

Web and Society; Web crawling; Human services; Non–profit organizations; Applied data science

ACM Reference Format:
Charalampos Chelmis and Mengfan Yao. 2019. Creating Public Value by Democratizing the Ecosystem of Human Service Providers. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3308560.3316487

1 INTRODUCTION

Navigating among service providers is an essential part of seeking for help [2, 7, 8, 11, 16, 20, 23, 25]. And yet, the process of seeking and providing services is often slow and fragmented: (i) service providers often work in silos [4, 10, 14, 15], whereas (ii) people who seek help carry the burden of navigating through available services and eligibility criteria [2]. In our ongoing project, we focus on the world of human service providers [12] with the ultimate goal of understanding how service seekers find the services they are looking for. Unlike the Web where people can find and access relevant information by reviewing websites or navigating webpages by following hyperlinks, no comprehensive resource exists to date for service seekers to find service providers. Additionally, the network structure (if any) among service providers is unknown. Thus, service seekers must rely on the information they already have (e.g., list of food pantries in an area) or new information they can obtain (i.e., through referrals) in order to eventually reach a service provider that can cater to their needs.

To understand how human service providers come together to address the grand challenges associated with meeting human needs, the most crucial problem to solve is finding an appropriate data source. To overcome this barrier, we leverage the medium that most (if not all) human service organizations rely upon for their public communication, i.e., their websites. The advantage of using websites to collect information about human service organizations is twofold. First, websites are publicly available online and their content is machine–readable. Second, they provide rich and semi–structured data that can be used to develop a comprehensive database of organizations and their partners, which can in turn be subsequently used to (i) develop technologies that enable novel capabilities, such as the one described in this paper, and (ii) analyze the interactions among human service organizations.

The limitation of using Web data is that it may lead to an imperfect representation of the actual set of human service organizations and their connections. For instance, organizations may choose to (i) not have a web presence at all, (ii) not list their partners on their websites, or (iii) use offline resources such as printed books similar to yellow pages to connect to other organizations. From the perspective of a service seeker however, information available on the Web may be the only resource to find services due to reasons, including but not limited to, restricted mobility (e.g., a homeless person may be unable to travel the distance to a referral organization so as to ask questions) or personal preferences (e.g., someone may prefer the anonymity of the Web to investigate for available options). Even though a subsequent study to understand the quality of this approximation may be necessary, it is beyond the scope of this work.

We begin by providing the key goal of our ongoing project (Section 2), and subsequently describe our data acquisition engine (and associated challenges) to collect the greatest amount of potentially useful information possible (including the mission and scope of an organization, hours of operation and list of services provided) about human service providers in a given locale (Section 4). We then illustrate the potential of the resulting repository of human service providers to create **public value** through a mobile app for

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3316487

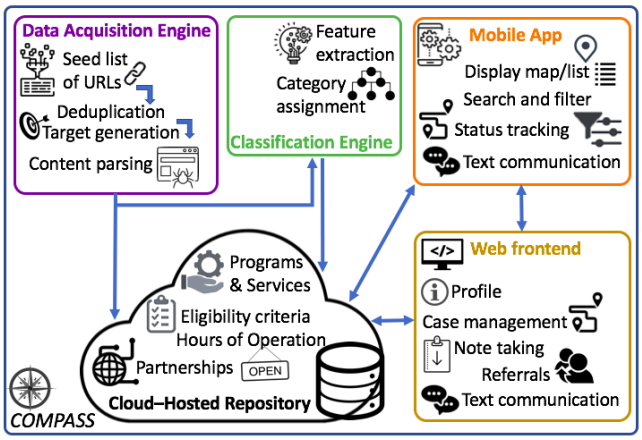


Figure 1: System architecture.

searching and filtering service providers, and requesting and tracking services through the click of a button (Section 5). We conclude with a discussion and future work in Section 6.

2 BACKGROUND

The principal objective of our ongoing project is threefold: (i) democratize access to the greatest amount of potentially useful information about available human services in a locale, without overwhelming the user with excess information, (ii) seamlessly connect service seekers to providers and facilitate digital interactions, and (iii) provide service seekers with a way to track the whole service cycle, from the moment a request is initiated, to service delivery, to outcome(s), all through a user–friendly, easy to use mobile app.

To accomplish this ambitious objective, we are currently developing *COMPASS*, a system comprising five modules: (i) data acquisition engine, (ii) classification engine, (iii) comprehensive repository and information integration engine, (iv) Web frontend, and (v) mobile app, as illustrated in Figure 1. In this paper, we present our progress on the data acquisition engine (Section 4) and the mobile app (Section 5) components of our framework.

3 RELATED WORK

Data collection from the Web. A large number of general–purpose crawlers have thus far been developed for scraping information from online sources [1, 13, 21]. While such crawlers are designed to collect data at scale, retrieving information about human services by scraping publicly available information from their websites presents unique challenges (see Section 4).

Technologies for Underserved Populations and Non–profit Organizations. Nearly 70% of low–income consumers rely primarily on their smartphones for connecting to the Web [9, 24]. Therefore, numerous mobile apps (e.g., Range[1] and OneToday[2]) have been developed to assist underserved populations get access to information, such as finding the nearest grocery store that accepts food stamps or the best prices for milk in a given area [3]. Apps specifically designed to address food insecurity and homelessness include [5, 6, 17, 18]. The use of technology in non–profit organizations has been explored in [27].

4 DATA ACQUISITION ENGINE

The goal of our data acquisition engine is to collect the greatest information possible about as many human service providers as possible within a given locale. In our current work, we focus on the metropolitan area surrounding Albany, the capital of the U.S. state of New York. Albany, the 4th largest metropolitan region in the state and the 45th largest in the U.S., is a multiracial and multiethnic city that contains nearly 99,000 residents (on average 52% White, 29% Black or African American, 9% Hispanic or Latino, 6% Asian, and growing refugee populations)[3] and a high concentration of human service providers and service organizations related to government, health care, and education [19]. The concentration of human service providers in Albany extends to two adjacent cities, Troy and Schenectady, which exist within a 20–mile radius from Albany. Albany and these two cities form the core of the Capital District of the New York State (officially defined as the Albany–Troy–Schenectady Metropolitan Statistical Area with a population size of approximately 800,000). Although these cities function as independent localities, they share commuting populations, local resources, and services. Nevertheless, our data acquisition engine is not restrained to the specific geographic boundaries of the Capital District, but is instead applicable to any location across the US.

We implemented our data acquisition engine in Python using BeautifulSoup[4] to parse HTML pages. A NoSQL cloud database[5] is used to store the data. The benefit of using a cloud–hosted database is security, persistency, flexibility, and real–time synchronization with every connected mobile app client.

4.1 Completeness

Ideally, a number of attributes, including the mission, physical address, hours of operation, programs and services provided, eligibility criteria, and category (as recorded by the National Taxonomy of Exempt Entities system,[6] which facilitates the categorization of organizations by their types and activities, promotes uniformity and comparability, and is used by the IRS to classify non–profit organizations), in addition to the website address and name of each organization, should be collected and stored in a comprehensive repository. Note that within the major group of "Human Services", the National Taxonomy of Exempt Entities system further divides organizations into logical subdivisions, the first few of which are shown in Figure 2.

Our data acquisition engine visits all webpages originating the starting domain of a given URL, parsing their content for the desired information. Note that we currently rely on the URL of an organization to identify it as unique; however, different URLs may resolve to the same webpage (e.g., due to parameters appended to the URL which do not change the content of the page). On the other hand, some organizations may use the same domain to host their

[1] http://www.rangeapp.org/
[2] https://onetoday.google.com/
[3] http://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2014/5-year.html
[4] https://www.crummy.com/software/BeautifulSoup/
[5] https://firebase.google.com/
[6] https://nccs.urban.org/classification/national-taxonomy-exempt-entities

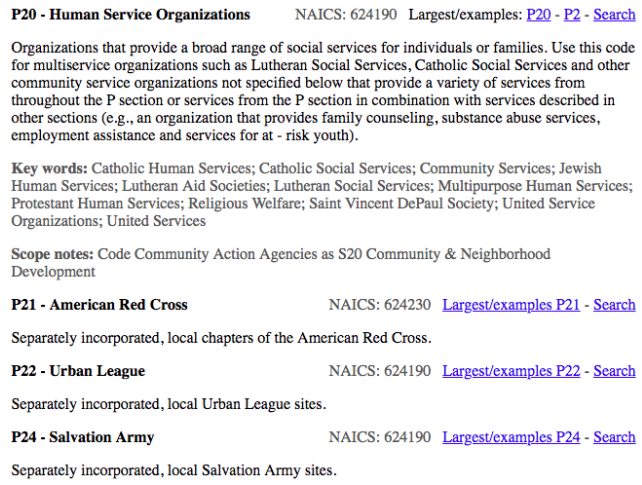


Figure 2: Sample subdivisions of human services according to the National Taxonomy of Exempt Entities system.

webpages. To address this challenge, we perform post–analysis to identify URLs resolving to the same organization. Finally, our data acquisition engine searches for partner organizations while ignoring links to social media platforms such as Facebook and Twitter so that no websites outside the realm of human service organizations are considered. If a domain has a subdomain, the subdomain is also examined.

Additionally, we would like our crawler to collect information about as many organizations in a region as possible. Our data acquisition engine performs snowball sampling, i.e., examines organizations that are listed as partners on websites of organizations in our seed list, to further expand the original list. The outcome depends to a large extent to the initial set of organizations provided as input to our data acquisition engine. Currently, our seed list comprises a total of over $3,000$ service–providing organizations and their corresponding URLs. This list was automatically compiled using Selenium[7] from (i) GuideStar,[8] an information service specializing in reporting on U.S. non–profit organizations, (ii) GreatNonprofits,[9] a website that allows donors, volunteers, and clients to share their personal experiences and reviews of charitable organizations, essentially providing crowdsourced information about the reputability of these organizations, and (iii) Charity Navigator,[10] an independent charity watchdog that evaluates charitable organizations in the United States.

4.2 Flexibility

Each organization's website has its own structure, which in addition may also change over time. Some organizations may have minimalistic websites containing a general overview of their mission and activities, while others might have multiple pages for specific programs and services. With respect to identifying webpages listing partnering organizations, these might appear in various formats, such as in menu items, images, PDF files, or unstructured text. The data acquisition engine should be able to detect and annotate the type of content describing partnering organizations, and either parse it automatically or log it and inform a system administrator to allow for post–processing and/or adaptation of the pertaining module. An easy solution would be to automatically filter navigation menus to identify pages that contain information about partner organizations. However, given differences and lack of standards in web design, this may be infeasible when crawling a number of heterogeneous websites.

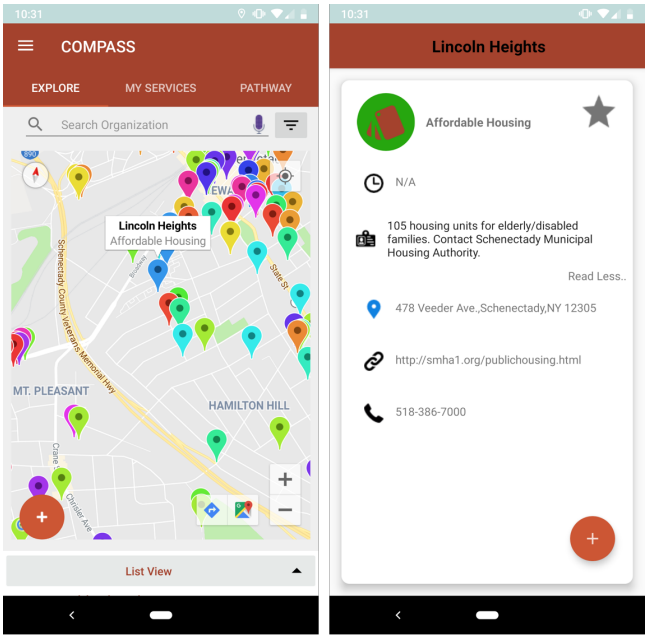
To address such difficulties, our data acquisition engine follows a modular design where its main functionality is implemented in generic modules (e.g., for the interaction with the cloud–hosted database), which do not require modification when adding new URLs for crawling. Our main content parsing module traverses a given domain searching for web pages containing information regarding the mission statement of organizations. Our partners detection module traverses a given domain searching for web pages containing at least three URLs with a ".org" suffix, as indicative of human service organizations, and therefore a potentially good heuristic to identify webpages that list partners. In our ongoing work, we are investigating the use of additional (and/or more appropriate) heuristics in order to enhance the ability of our engine to parse additional formats so as to better capture existing relationships between organizations. When a website uses a previously unseen structure, a new scraping module (i.e., set of regular expression matching operations to identify specific patterns) may be required.

4.3 Efficiency

Information about each parsed organization should be stored in our comprehensive repository exactly once, and the same organization should not be visited more than once during a crawl. In our current implementation, we perform a basic edit distance string similarity check on organizations' title in addition to examining their corresponding URLs to identify and avoid parsing and storing duplicate entries. Our engine additionally provides the means to revisit a website, inspect it, and if needed, update the entry in our repository with new or modified information about the programs and services and partnering organizations of the corresponding service provider. A log system helps to keep track of the crawling status and detect unexpected errors (e.g., site is down) or encountered problems (e.g., network connectivity issues). Most errors and warnings can be handled automatically (e.g., connectivity problems), however some issues require manual inspection (e.g., when a website uses a previously unseen structure). Self–recovery capabilities are used to detect when an error has occurred, and if possible, continue with the data collection process.

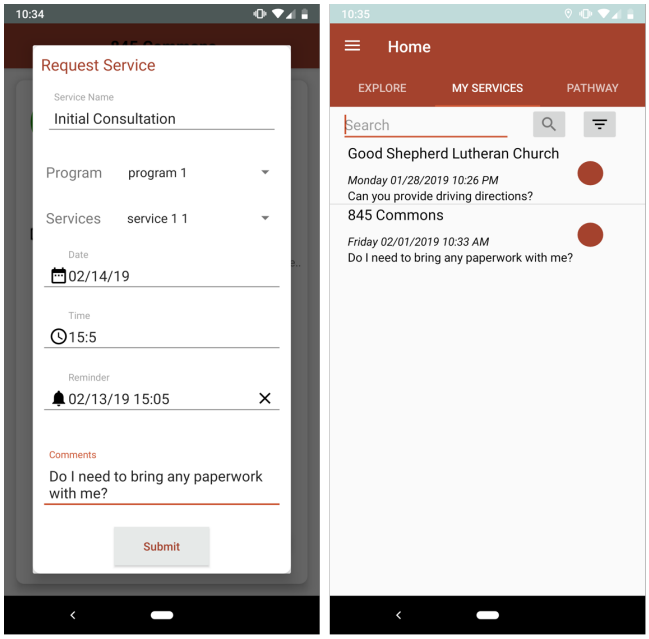
Although possible to set up our data acquisition engine to be launched in parallel using different processes and/or proxies (e.g., to parse multiple URLs in parallel), we decided against such option for two main reasons. First, we wanted to balance the use of our available bandwidth, and second, we did not want to compromise or cause any harm whatsoever to the normal operation of any of the websites visited during the data acquisition phase. Note that scrapping data may already require to break some of the terms of

[7] https://www.seleniumhq.org/
[8] https://www.guidestar.org/
[9] https://greatnonprofits.org/
[10] https://www.charitynavigator.org/



(a) (b)

Figure 3: Screenshots of our application prototype illustrating: (a) map view of human service providers color-coded by category, and (b) details about a selected provider.



(a) (b)

Figure 4: Screenshots of our application prototype showing (a) a user–friendly and easy to use dialog where the user provides inputs regarding her request to the provider, and (b) status page, where the user is able to keep track of her service requests (colors are used to differentiate between pending, cancelled, under processing, or completed requests).

use agreements used by human service providers' websites. We believe however that the benefits of crawling for data collection outweigh the potential harms.

Finally, the data acquisition engine is useful to the extent it can correctly identify the category a newly found human service provider belongs to for a large percentage of organizations based on textual information (e.g., mission statement, and available programs and services) and the category of known partnering organizations already in our repository. However, automatic categorization of organizations is a challenging problem itself. Currently, we (i) cross–reference information from GuideStar, GreatNonprofits, and Charity Navigator to assign a category to organizations already in our database, and (ii) manually annotate organizations beyond the initial list. We plan to deploy a machine learning approach to automatically perform this task as part of our classification engine module in our future work.

5 ENABLING HUMAN SERVICES 2.0

The term "Web 2.0" refers to a " *set of economic, social, and technology trends that collectively form the basis for the next generation of the Internet, a more mature, distinctive medium characterized by user participation, openness, and network effects*" [22], whereas, terms "Health 2.0" and "Medicine 2.0" have been recently used to describe the use of "Web 2.0" technologies in health care [26]. In similar spirit, we envision "Human Services 2.0", the next generation of human services, in which (i) information about service providers is openly accessible and easy to find, search and filter through, and (ii) service delivery is digital, seamless, worry–free, efficient, and traceable, maximizing positive outcomes both for service seekers and service providers alike.

To realize the potential of "Human Services 2.0", we leverage the extensive repository created by our data acquisition engine (Section 4) to open up information about human services and facilitate the digital connectivity with human service providers across conventional boundaries and silos. In our proof–of–concept prototype, a service seeker opens the application, and is prompted with a map view of human service providers nearest to her current geographic position (Figure 3a), which is provided either by allowing the application to gather her current GPS location or by dragging a pin marker to the appropriate location. A user can additionally explore human service providers in the form of a list (similar to an offline directory), or expand on any provider to obtain more information, including directions and contact details, hours of operation, programs and services provided, and eligibility criteria if applicable (Figure 3b). Filtering and search capabilities (Figure 5), and the ability to bookmark organizations (star symbol at the top right corner of Figure 3b) enable users to easily navigate among the potentially large number of service providers, whereas requesting for services become as easy as the click of a button (illustrated in Figure 4a). Tracking the status of service requests is simplified as shown in Figure 4b, whereas notifications can be set (some automatically by the app) to notify the user of changes in any of her requests (e.g., need for additional documentation). Our mobile app additionally

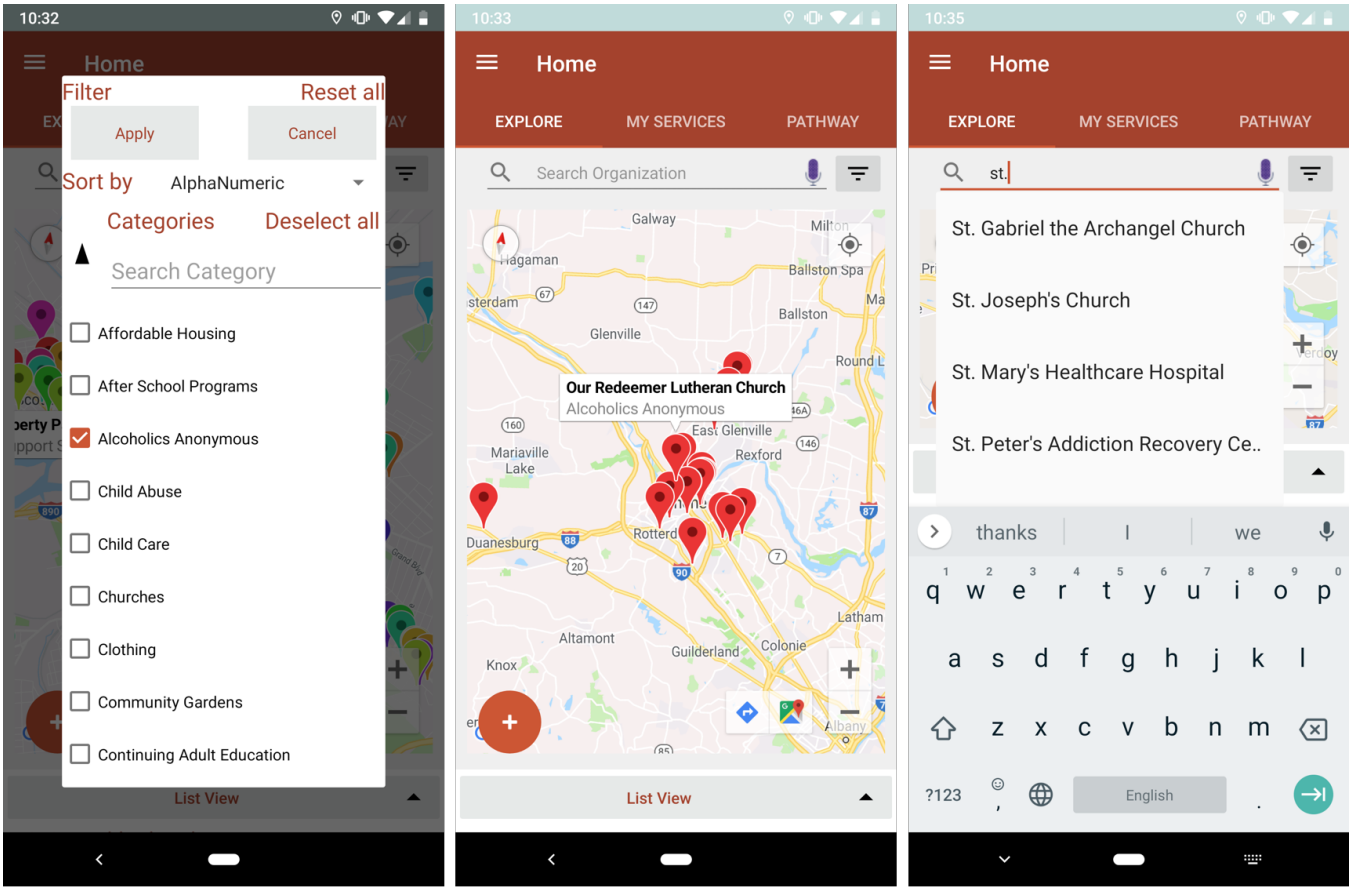


Figure 5: Screenshots of our application prototype showing (a) filtering and (c) search capabilities to narrow down one's search to a small subset of providers, as in (b) where only providers offering "Alcoholic Anonymous" services are displayed.

supports asynchronous text communication between the user and service providers.

6 CONCLUSION

The promise of *COMPASS* lies in its ability to collect and filter useful information about human service providers from a mass of semi–structured data, and democratize access to such data through a user–friendly mobile app. While still in early stages of development, *COMPASS* has generated significant interest from (i) the United Way of the Greater Capital Region[11], a non–profit organization that brings together non–profits, businesses, government and human service agencies, schools, organized labor, financial institutions, community development corporations, voluntary and neighborhood associations, and the faith community to address basic needs, education, income and health issues in the Capital Region, (ii) the Community & Public Service Program at the University at Albany, State University of New York[12], a University–wide initiative that pairs School of Social Welfare students with local non–profit and public organizations to collaboratively address issues that impact the well–being of marginalized populations, and (iii) ServeAlbany[13], a grassroots initiative with the sole goal of simplifying the process for those willing to volunteer their time and services to "*find a place to get involved and make a difference*".

Nevertheless, a number of improvements remain to be made for it to become a truly useful resource for both the general public, human service providers, and researchers. For example, we are exploring the use of additional sources to improve our data acquisition process. In pursuit of automatic categorization of newly discovered human service providers, we are exploring state–of–the–art natural language processing techniques, clustering methods, and network science principles to produce plausible labels as part of a semi–supervised or semi–automated process. With respect to the mobile app, we are working towards improving the user experience with added features such as keyword searching and improve customization such as personalized recommendations and notifications. We also plan to conduct a usability observation study to gather feedback from our target demographic on priority features

[11] https://www.unitedwaygcr.org/
[12] https://www.albany.edu/cpsp/
[13] https://servealbany.org/

as well as how to best improve the mobile app user–experience. Beyond user–evaluation, we are working towards developing quantitative evaluation metrics for the comprehensive system, including its ability to limit noise (i.e., organizations beyond the human service domain), correct classification of organizations and coverage as compared to manually curated sources such as GuideStar and GreatNonprofits.

7 ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. ECCS-1737443. We also want to thank Mr. Balaji Sankeerth Jagini for his initial contribution to the development of the data acquisition engine, Mr. Dhruvil Kotak and Bhavik Patel for their contribution towards developing our mobile app prototype, and Drs. Daphney–Stavroula Zois and Wonhyung Lee for their constructive feedback that helped improve the functionality of our mobile app prototype.

REFERENCES

[1] Paolo Boldi, Bruno Codenotti, Massimo Santini, and Sebastiano Vigna. 2004. Ubicrawler: A scalable fully distributed web crawler. *Software: Practice and Experience* 34, 8 (2004), 711–726.

[2] Judith B Bradford, Sharon Coleman, and William Cunningham. 2007. HIV System Navigation: an emerging model to improve HIV care access. *AIDS patient care and STDs* 21, S1 (2007), S–49.

[3] Deana Brown and Rebecca E Grinter. 2016. Designing for transient use: A human-in-the-loop translation platform for refugees. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 321–330.

[4] Tom Corbett and Jennifer Noyes. 2008. *Human services systems integration: A conceptual framework*. Institute for Research on Poverty.

[5] Tawanna R Dillahunt and Amelia R Malone. 2015. The promise of the sharing economy among disadvantaged communities. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2285–2294.

[6] Lynn Dombrowski, Jed R Brubaker, Sen H Hirano, Melissa Mazmanian, and Gillian R Hayes. 2013. It takes a network to get dinner: designing location-based systems to address local food needs. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 519–528.

[7] Daniel Eisenberg, Ezra Golberstein, and Sarah E Gollust. 2007. Help-seeking and access to mental health care in a university student population. *Medical care* (2007), 594–601.

[8] Chris Ferguson. 2000. Shaking the conceptual foundations, too: Integrating research and technology support for the next generation of information service. *College & Research Libraries* 61, 4 (2000), 300–311.

[9] Thom Fil and Camille Ryan. 2014. *Computer and internet use in the United States*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

[10] Patricia Fredericksen and Rosanne London. 2000. Disconnect in the hollow state: The pivotal role of organizational capacity in community-based development organizations. *Public Administration Review* 60, 3 (2000), 230–239.

[11] Nancy Gourash. 1978. Help-seeking: A review of the literature. *American journal of community psychology* 6, 5 (1978), 413–423.

[12] Yeheskel Hasenfeld. 2009. *Human services as complex organizations*. Sage Publications.

[13] Allan Heydon and Marc Najork. 1999. Mercator: A scalable, extensible web crawler. *World Wide Web* 2, 4 (1999), 219–229.

[14] Marcela Horvitz-Lennon, Amy M Kilbourne, and Harold Alan Pincus. 2006. From silos to bridges: Meeting the general health care needs of adults with severe mental illnesses. *Health affairs* 25, 3 (2006), 659–669.

[15] Kenneth Kernaghan. 2005. Moving towards the virtual state: Integrating services and service channels for citizen-centred delivery. *International Review of Administrative Sciences* 71, 1 (2005), 119–131.

[16] Satya P Krishnan, Judith C Hilbert, and Dawn VanLeeuwen. 2001. Domestic violence and help-seeking behaviors among rural women: Results from a shelter-based study. *Family & Community Health* 24, 1 (2001), 28–38.

[17] Christopher A Le Dantec, Robert G Farrell, Jim E Christensen, Mark Bailey, Jason B Ellis, Wendy A Kellogg, and W Keith Edwards. 2011. Publics in practice: Ubiquitous computing at a shelter for homeless mothers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1687–1696.

[18] Min Kyung Lee, Ji Tae Kim, and Leah Lizarondo. 2017. A Human-Centered Approach to Algorithmic Services: Considerations for Fair and Motivating Smart Community Service Management that Allocates Donations to Non-Profit Organizations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3365–3376.

[19] Wonhyung Lee. 2017. Sustainability of nonprofit human service organizations in a neighborhood context. *Nonprofit Management and Leadership* 28, 1 (2017), 11–24.

[20] Ramin Mojtabai, Mark Olfson, and David Mechanic. 2002. Perceived need and help-seeking in adults with mood, anxiety, or substance use disorders. *Archives of general psychiatry* 59, 1 (2002), 77–84.

[21] Christopher Olston, Marc Najork, et al. 2010. Web crawling. *Foundations and Trends® in Information Retrieval* 4, 3 (2010), 175–246.

[22] Tim O'reilly. 2009. *What is web 2.0*. O'Reilly Media, Inc.

[23] Emmanuel Scheppers, Els Van Dongen, Jos Dekker, Jan Geertzen, and Joost Dekker. 2006. Potential barriers to the use of health services among ethnic minorities: A review. *Family practice* 23, 3 (2006), 325–348.

[24] Aaron Smith and Dana Page. 2015. US smartphone use in 2015. *Pew Research Center* 1 (2015).

[25] Michael Ungar. 2005. Pathways to resilience among children in child welfare, corrections, mental health and educational settings: Navigation and negotiation. In *Child and youth care forum*, Vol. 34. Springer, 423–444.

[26] Tom H Van De Belt, Lucien JLPG Engelen, Sivera AA Berben, and Lisette Schoonhoven. 2010. Definition of Health 2.0 and Medicine 2.0: a systematic review. *Journal of medical Internet research* 12, 2 (2010).

[27] Amy Voida, Ellie Harmon, and Ban Al-Ani. 2012. Bridging between organizations and the public: volunteer coordinators' uneasy relationship with social computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1967–1976.