"Enlightening, Inclusive, Inspiring": iConference 2019 Conference Review Post-print

Authors: Gonghong Village, Shen Yutian, Cao Gaohui

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] iConference is an important international conference in the field of information science. This article aims to summarize and systematically review the papers from iConference 2019, forming a relatively comprehensive review to provide reference and guidance for domestic researchers in this field. [Method/Process] From four dimensions—research methods, research questions, research process, and research findings—this study reviews 77 papers covering 24 themes accepted by iConference 2019. [Results/Conclusion] As the annual conference of iSchools, the themes of accepted papers encompass the latest achievements in theoretical research and practical applications in the current field of information science. The results indicate that the data-driven research paradigm provides opportunities for the integrated development of information science and other disciplines, prompting library and information science to identify research questions in new research contexts. The combination of qualitative and quantitative methods, such as semi-structured interviews, questionnaires, grounded theory, and statistical analysis, constitutes the current primary research methods. The acquisition of research data is an extremely important process in research.

Full Text

Preamble

Gong Hongcun[1], **Shen Yutian**[1], **Cao Gaohui**[1,2]
[1]School of Information Management, Central China Normal University, Wuhan 430079
[2]Center for Data Governance and Intelligent Decision-Making Research, Hubei Province, Wuhan 430079

Abstract

[**Purpose/Significance**] iConference is an important international conference in the field of information science. This paper aims to summarize and review the papers presented at iConference 2019, providing a comprehensive overview for domestic researchers in this field as a reference. [**Method/Process**] This review examines 77 papers across 24 themes from iConference 2019 through four dimensions: research methodology, research questions, research process, and research findings. [**Result/Conclusion**] As the annual conference of iSchools, iConference covers the latest achievements in theoretical research and practical applications in information science. The results demonstrate that the data-driven research paradigm provides opportunities for the integration and development of information science with other disciplines, prompting libraries and information disciplines to discover research questions in new research scenarios. Combined qualitative and quantitative methods such as semi-structured interviews, questionnaires, grounded theory, and statistical analysis are currently the main research approaches, and research data acquisition is an extremely important process in these studies.

Keywords: iConference; conference review; data-driven; information science

---

1. Conference Introduction

The 2019 (14th) iConference was held from March 31 to April 3, 2019, at the University of Maryland, College Park. Hosted by the University of Maryland, College Park, and co-hosted by Syracuse University and the University of Maryland, Baltimore County, the conference received sponsorship from numerous institutions including the National Science Foundation, Computing Research Association, Emerald Publishing, Elsevier, MDPI, Morgan & Claypool Publishers, NVIVO, the University of Pittsburgh School of Computing and Information, the University of Kentucky College of Communication and Information, the National Taiwan University Department of Library and Information Science, and the University of Tennessee, Knoxville School of Information Sciences. The conference theme, "Inform, Include, Inspire," aimed to explore the meaning of being informed in the 21st century, how to expand the reach of the information revolution, and how to better inspire individuals and organizations to utilize information in a rapidly changing knowledge society. Over 60 deans from iSchools worldwide and nearly 600 scholars attended the event. A total of 77 papers were accepted, including 33 Short Papers (acceptance rate 37.5%) and 44 Full Papers (acceptance rate 33%). The accepted papers were published in the LNCS (Springer's Lecture Notes in Computer Science) proceedings [1], with posters and selected papers made openly accessible through the University of Illinois' IDEALS (Illinois Digital Environment for Access to Learning and Scholarship) platform [2]. The following review is based on these 77 conference papers, which were organized into 24 research themes.

2. Conference Content

2.1 Scientific Work and Data Practices

Data-intensive scientific discovery, as the fourth paradigm of scientific research, has rapidly developed across multiple disciplines in both social sciences and natural sciences. Efficient sharing, utilization, and management of massive scientific data constitute an important foundation for data-driven scientific research. Youngseek Kim from the University of Kentucky constructed a theoretical framework for research data sharing based on institutional theory and the theory of planned behavior, using questionnaire surveys to collect data and examine how institutional pressures, resources, and individual motivations influence bioscientists' open sharing of different types of research data (genomic sequence data and laboratory data). The study found that these three factors—institutional pressure (requirements from funding agencies and academic journals), resources (metadata standards and data repositories), and individual motivations (perceived career benefits, perceived career risks, individual effort)—have significantly different effects on bioscientists' open sharing of different data types, suggesting that different policy measures should be adopted to encourage data sharing. Drew Paine and Lavanaya Ramakrishnan from Lawrence Berkeley National Laboratory conducted semi-structured interviews with five data producers and five data users from astronomy and earth sciences to understand various aspects of their research projects, focusing on their data usage and acquisition methods. They also inquired about how researchers determine versions of data products for analysis and the impacts (both expected and unexpected) of data product changes. Using an improved grounded theory approach to code and analyze interview data, they explored how scientists perceive data changes in their research work, aiming to design tools and best practices to support scientists in developing and maintaining their research infrastructure and to help them examine data changes to discover critical information embedded within their research infrastructure.

2.2 Methodological Concerns in (Big) Data Research

The data-driven research paradigm has promoted the development of modern scientific research while simultaneously facing challenges such as insufficient methodological standardization, poor interpretability of results, and complexity in data analysis and visualization. Addressing methodological issues in big data research, Tingting Jiang from Wuhan University used grounded theory to extract 279 information encountering events from 14 research papers on the topic, forming nine major categories, 33 subcategories, and 230 conceptual labels to abstractly construct a model of the information encountering process and analyze relationships among major elements in detail. Madelyn Rose Sanfilippo from Princeton University and Chase McCoy from Indiana University selected the 50 most-cited papers from 2010-2018 in three fields—"Government

and Law," "Education Evaluation," and "Information and Library Science"—from Web of Science (including only empirical studies with clearly outlined methodologies). Using meta-analysis, they compared and evaluated differences between data-driven and statistical analysis research methods. Results showed that traditional statistical analysis methods are more institutionalized and consistent, while data-driven approaches have certain limitations in reliability, validity, generalizability, and interpretability. Jaime Snyder from the University of Washington and Katie Shilton from the University of Maryland compared two visualization tools (R and Adobe Illustrator) from the perspectives of audience, support structure, licensing, usage and scalability, graphic capabilities, data input/import methods, graphic output, design patterns, aesthetics, layout, and optimization features. They extracted descriptions of software features and functions over time from software documentation (release notes, user guides, etc.), training materials (online tutorials, videos, demonstrations, etc.), and marketing/media materials (new feature announcements, advertisements, reviews, blog posts, etc.) to track product updates and changes. They also compared and evaluated the value of these tools in terms of professional terminology and functionality, further discussing the significance of data-driven work and data science curricula in the current environment. They concluded that data visualization practices are converging in the current environment, while professional terminology in software product introductions remains divergent.

2.3 Concerns About "Smart" Interactions and Privacy

With the development of Internet of Things technology, smart devices and smart homes are widely applied in various life scenarios, bringing convenience while also raising user concerns about privacy leakage. The universality of technology and the degree of privacy protection affect user acceptance of smart devices. Toini Bogers from Aalborg University Copenhagen used questionnaire surveys to investigate how 357 Danish-speaking users utilize smart personal assistants. Results showed that only 19.9% of users employed smart personal assistants, with most considering their functions unreliable. A study on Siri's functional availability found that when interacting in Danish, Siri exhibited poor speech recognition, unnatural dialogue responses, and inability to support mixed-language speech recognition. Chola Chhetri and Vivian Genaro Motti from George Mason University scraped 66,656 user reviews of Amazon Echo, Google Home, Samsung SmartThings Hub, Wink Hub, and Insteon Hub from Amazon.com and Bestbuy.com. They manually coded reviews across four dimensions: attention level (specific vs. non-specific), emotional polarity (positive, neutral, negative), temporal sequence (collection, transmission, storage, sharing), and security principles (confidentiality, integrity, availability, authenticity) to analyze user privacy concerns regarding smart home devices, proposing functional recommendations to enhance privacy protection in smart home products. Yuting Liao from the University of Maryland and Michael Zimmer from the University of Wisconsin-Milwaukee used online surveys to collect 1,160 sample data points (including both smart personal assistant users and non-users). Based on

social contract theory and the technology acceptance model, they explored how user motivations and barriers, as well as trust concerns about data privacy and corporate compliance with IPA-related agreements, affect IPA acceptance and usage.

2.4 Identity Questions in Online Communities

Users in online communities construct a virtual identity for themselves through various online behaviors (posting information, purchasing virtual equipment, setting user avatars, etc.). Simultaneously, online communities attract users to actively build virtual identities through mechanisms like point systems and ranking systems. Identity recognition is an important factor influencing some users' continued community usage. Travis Faas from Indiana University used semi-structured interviews to investigate eight Twitch streaming platform users with 10 or fewer viewers, employing open coding for thematic analysis of interview content to study the motivations of streamers with small audiences. The study found that half of the interviewed users were in transitional stages regarding work, city, or lifestyle. Their motivations for continued streaming were twofold: first, work needs, as they could obtain audience opinions and help regarding their work during live broadcasts; second, reducing loneliness, as streaming allowed them to meet friends they couldn't encounter in real life. Based on these findings, the authors recommended that streaming websites design live communities and learning environments for users in life transition periods to meet learning needs of both users and audiences and help users seeking assistance find people who can help them. Alia Reza from Stony Brook University investigated literature related to representation characteristics of game characters in microtransactions and purchasing behaviors related to players' racial identities, studying how character representation in microtransactions affects purchasing behaviors of players from different racial backgrounds. The study found that purchasing skins through microtransactions allows players to choose more diverse appearances than default skins, enabling greater self-expression and enhancing possibilities for social belonging. However, the commercialization of character appearances (such as skin color) may also increase psychological burden for players of certain races. Kathryn E. Ringland from Northwestern University used virtual ethnographic methods to collect approximately 200 hours of immersive observation data from Autcraft users between May 2014 and May 2017, including about 5,000 forum topics and over 150 blog posts. Using an iterative and inductive approach to identify, analyze, and name phenomena in the data, the study found that the label "autism" is not only a source of disenfranchisement but also a source of empowerment and identity. The author argued that online communities can provide relatively closed and safe spaces for individuals with autism but cannot serve as medical treatment approaches; while identifying as autistic in the community can provide a sense of identity, it is also important to encourage autistic individuals to bravely face the real world.

2.5 Measuring and Tracking Scientific Literature

Dong Joon Lee from Texas A&M University extracted relevant data for engineering faculty from Texas A&M University from Scholars@TAMU and categorized these faculty into three groups based on awards and honors data collected by the National Research Council (NRC) [3]: those receiving high prestige (HP) awards, those receiving prestigious (P) awards, and those receiving no awards. Using Kruskal-Wallis tests to explore relationships between faculty research performance indicators and research collaboration practices, and Bonferroni tests to determine the relative importance of performance indicators across categories, results showed that HP award recipients had significantly higher average rankings than non-award recipients in total citations, citation counts of cited articles, h-index, and total publications. HP award recipients also had significantly higher average rankings than P award recipients in citation counts and total publications, but no significant differences existed between P award recipients and non-award recipients. University administrators and bibliometric researchers can use these findings to combine award attainment with other research performance indicators and develop machine learning models to predict whether faculty are likely to receive HP awards. Michael Lesk from Rutgers University collected articles from PLoS between 2014-2018 and their 153 retracted papers (via Retraction Watch), using statistical analysis to study the relationship between open data and trustworthy research. Results showed that approximately half of PLoS papers had accompanying data, with the proportion of data availability in retracted papers below average—about 27% of retracted papers had open data. The study concluded that papers with high-quality, well-curated open data are more reliable, should receive higher credibility, deserve greater prestige within university communities, and should obtain higher rankings in search engines. Ania Korsunska from Syracuse University used Google search engine as a data source, tracking citations to specific articles through hyperlink references to explore the spread of scientific misinformation. Specific questions included: How does misinformation flow between different media sources? Which media channels rely on primary sources versus secondary sources? Which media channel is the source of misinformation? Results showed that misinformation spreads faster and more widely than facts, mass media over-relies on secondary sources, publication reputation affects public credibility of information, and popular media publications serve as information sources for each other. Tong Zeng from Nanjing University used 1.9 million article URLs from PubMed Open Access to explore the reusability of shared resources in open access publications. Using logistic regression to study how URL-based resource sharing evolves over time and factors affecting resource availability, results showed that URLs have become the primary mechanism for sharing resources, with half of links becoming unavailable after eight years, though this trend does not continue in the second eight-year period. URL decay does not follow the typical half-life analysis found in physics research; resource usage frequency, international and organizational domain suffixes, and the number of affiliations in papers positively correlate with link availability, while link age, length, and domain suffixes for India (in),

EU (eu), and China (cn) negatively correlate with availability.

2.6 Limits and Affordances of Automation

Felix Hamborg from the University of Konstanz employed established models and methods from social sciences to describe and analyze media bias, while utilizing scalable text analysis methods from computer science as technical support, proposing a system prototype (Newsalyze). The prototype identifies bias instances by selecting words and labels (WCL) from a group of news articles reporting on the same event, using U.S. DACA and Denuclearization and UK Brexit news to validate the model's effectiveness. Emi Ishita from Kyushu University proposed a three-stage automated content analysis process: determining documents to be labeled at the document granularity level, determining which text within documents should be labeled at the sentence granularity level, and assigning labels that reflect human values to this text. They collected news reports related to nuclear energy and power plants from the 2011 Fukushima nuclear leak through 2016, with two researchers conducting 10 rounds of annotation to obtain a corpus of 448 reports. They then used 348 reports as a training set to train two classifiers, TinySVM and fastText, and the remaining 100 reports as a test set. Additionally, two researchers annotated valuable sentences in 28 reports, evaluating classifier performance by comparing precision, recall, and F-values, with results showing good classifier performance. Matt Willis from the Oxford Internet Institute used field surveys to observe and conduct semi-structured interviews with staff at six primary care practices across England, collecting relevant documents from general practitioner clinical practices to study challenges in integrating AI into automated GP clinical documentation practices. The survey found that GP documentation practices have high timeliness requirements; notes in medical records vary from simple and quick to complex and time-consuming; GPs demonstrate personal preferences in interpersonal style and documentation methods; and smart speakers and other intelligent devices can improve documentation efficiency. The study showed that automation can effectively prevent GP burnout and stress, but due to implicit information and environmental influences in documentation processes, AI cannot fully automate complex tasks. Therefore, the authors recommended collaborative work between medical personnel and AI systems for documentation practices.

2.7 Collecting Data about Vulnerable Populations

Research on information needs and behaviors of vulnerable populations (impoverished individuals, immigrants, persons with disabilities, those with language barriers, etc.) helps public information service institutions provide effective services. Jordan Dodson from the University of North Carolina at Chapel Hill used questionnaire surveys to investigate health self-tracking behaviors and digital tool usage for depression management among 83 first-generation (immigrated to the U.S. after age 17) to 2.5-generation (born in the U.S. but with one parent not

born in the U.S.) immigrant college students with depression. Results showed that 94% of students with depression engaged in health self-tracking, but very few tracked mental health indicators, with 81.9% not using APP tools and 86.7% not using other online resources to help treat depression. The study raised questions about why health digital tools are not used and how they might be utilized. Safirotu Khoir from Gadjah Mada University in Indonesia analyzed literature in the information behavior field using the Photovoice research method, exploring from a methodological perspective the superiority and feasibility of Photovoice in investigating immigrants' information needs. Photovoice is a method using photos to convey information, effectively bridging cultural differences and language barriers between researchers and participants and helping capture participants' thoughts and emotions. The study demonstrates the value of photos in researching vulnerable groups and non-English-speaking populations. Devendra Potnis from the University of Tennessee and Bhakti Gala from Central University of Gujarat developed a cross-disciplinary six-step tool for the poor to improve mobile applications and financial literacy, including questionnaires, focus group prompts, and practical training. The study provides references for public libraries and governments in developing countries to promote mobile financial transactions for impoverished populations. Sara Vannini from the University of Washington used group interviews to survey information privacy awareness and practices among nine humanitarian organization personnel working with undocumented immigrants in the U.S., using qualitative analysis software Dedoose to identify thematic areas and periodic themes. Results showed that systems used by humanitarian organizations to collect and store information about vulnerable groups face technical and human factor risks; untrained staff and volunteers may unintentionally leak sensitive information; and organizations lack oversight for information leakage issues. Based on these findings, the authors proposed Humanitarian Information Activity Privacy (HIAP) guidelines to ensure adequate privacy protection for vulnerable groups in irregular migration contexts.

2.8 Supporting Communities Through Public Libraries and Infrastructure

Public libraries and other information infrastructure built by governments are valuable for meeting community information and cultural needs and constitute important components of community building. Kathleen Campana from Kent State University and J. Elizabeth Mills and Michelle H. Martin from the University of Washington used focus groups, interviews, and national surveys to collect data from 157 library staff at small, medium, and large public libraries in urban, suburban, and rural areas. The study explored how libraries collect and utilize data on the needs of families and children in underserved communities to serve community members more effectively. Results showed that most libraries actively use various methods to collect data, with conversational methods being primary, but fewer using structured, sustained methods like focus groups and interviews. Communities (including community members, collaborators, leaders, etc.) are the most common data sources for libraries, but libraries are also

seeking other types of data on family needs, such as census data. Most libraries (76%) consciously use this needs data for service development, but only some (35%) do so continuously. J. Elizabeth Mills from the University of Washington and Kathleen Campana from Kent State University built upon the VIEWS2 project research, constructing the VIEWS2 conceptual framework from three building blocks—intentionality, interactivity, and community—to support the design and production of early learning programs in libraries, interviewing experimental group participants. Interview results showed that the framework's building blocks may influence storytime providers' practices. Specifically, intentionality in planning early literacy content for storytime increased due to intervention (the VIEWS2 conceptual framework); content interactivity helped them expand and explore new activities; and interaction with colleagues reduced loneliness and provided experience and work feedback. Sang Hoo Oh and Marcia A. Mardis from Florida State University examined the extent to which rural broadband in the U.S. can support advanced manufacturing development through literature analysis, specifically studying the importance of advanced manufacturing to rural development, rural broadband conditions, the importance of rural broadband internet to advanced manufacturing, and related policy documents. Results showed that while various initiatives and programs have been implemented to support advanced manufacturing development and broadband services in rural areas, some rural regions still lack broadband access and receive poor quality broadband internet services. Rural communities can benefit from the influx of high-tech innovation industries, but it remains unclear whether these communities can provide adequate broadband support. Colin Rhinesmith from Simmons University used participatory research methods to work with research groups at ten U.S. public libraries to co-design a broadband measurement platform and training manual, proposing an information system design solution for an open-source broadband measurement platform for U.S. public libraries to help them better understand the relationship between their network infrastructure and digital services and to inform local, state, and national broadband planning efforts.

2.9 Information Behaviors in Academic Environments

College students represent a group with relatively high information literacy, diverse information needs, and high information usage frequency. Research on their information behaviors has important application value, and because college student samples are relatively easy to obtain, such research has always been an important topic in information behavior studies. Vincent Grimaldi from Suffolk University used semi-structured interviews to investigate 15 student-athletes at three U.S. institutions, examining the role of Information and Communication Technology (ICT) in college athletes' information seeking, daily behaviors, and long-term goal setting during their transition from students to graduates. The study found that ICT can improve college athletes' information seeking capabilities and access efficiency; using ICT has become a daily behavior for college athletes, enhancing muscle memory, optimizing game strategies, and enabling

targeted training; and ICT plays a positive role in helping college students formulate long-term professional development goals. Kyong Eun Oh and Rong Tang from Simmons University collected diary entries over three weeks from 51 U.S. college students about their mobile phone news consumption using a diary template in a mobile application (each entry included participants' evaluations and emotions about news). Using a diary study method, they investigated how college students browse news on mobile phones and their reactions to inclusive/exclusive political and social news, finding that college students prefer inclusive news and reject exclusive or unfair news, indicating that U.S. college students have sensitive awareness of social justice and political issues.

2.10 Data-Driven Storytelling and Modeling

Storytelling has become an important research topic in information science in recent years, representing an event-centered approach to information organization and presentation. Xiying Wang from the University of Pennsylvania developed a Web application called Memofest to collect festival stories from participants in a local arts festival, using semi-structured interviews with 18 participants and coding the results. Four themes were identified from festival stories: festival activities, artworks, social activities, and recollections of place changes. Three types of participants with parallel experiences were identified from interview content: experience explorers, nostalgic travelers, and familiar explorers. The study provided recommendations on how information technology can promote local development for these three participant types. Zhenpeng Zhao from the University of Maryland used data comics as research objects to explore the importance of two narrative mechanisms in comic storytelling—partitioning and sequencing (ordering of frames to create narrative). Using an experimental method, participants answered questions using complex infographics and compared this with answering questions using data comics organized from the same infographics, rating them on attractiveness, speed, spatial efficiency, ease of use, and interest. Results showed that data comics provide a more effective storytelling method than single visualizations, with partitioning and sequencing mechanisms helping participants focus on and recall stories. Enguo Zhou from Sun Yat-sen University proposed a new model—ABGM (Adoption Behavior-based Graphical Model)—to deeply study the innovation diffusion process. The model defines two types of relationships: homogeneity and heterogeneity, and organically connects social networks, innovation networks, and correlation networks within a unified framework. The model was tested on a large-scale artificial intelligence publication dataset from 2006-2015, with experimental results showing that compared with three baselines (logistic regression, support vector machine, Adamic/Adar), homogeneity and heterogeneity in adoption behavior provide relevant information for predicting adoption decisions, and the probability of individual users adopting innovations is significantly influenced by correlation network diffusion processes.

2.11 Online Activism

Shipi Dhanorkar and Mary Beth Rosson from Pennsylvania State University examined online petition platforms from a crowdsourcing perspective, viewing them as a special type of crowdsourcing system specifically designed to promote social change. Using purposive sampling, they selected four online petition platforms—Change.org, Care2petitions, MoveOn.org, and Wethepeople—and conducted comparative analysis across four dimensions: petition initiation, petition updates, petition signing, and petition analysis. Based on this analysis, they proposed design improvements for online petition platforms: providing participation motivation, establishing reputation for petition initiators, achieving critical mass for collective action, and message framing to balance the "openness" and "quality" of online petitions. Jennifer Pierre from the University of California used mixed methods to conduct content analysis of social media site posts on five Facebook pages from four aspects: overall situation, liking, commenting, and sharing behaviors, media, and themes, focusing on original article content and related comments and activities. The study aimed to examine the popularity of content expressing or intending to evoke "kamamuta" and explore kamamuta's role in stimulating individual positions and broader social participation in social media activism. Results showed that kamamuta-inducing content would be more widely shared and liked. Pu Yan from Oxford University used case study methods, utilizing the Weibo social platform to conduct visual analysis of dialogue data from five government/media accounts and five intellectual/online activist accounts during the "February 2014 Dongguan anti-prostitution campaign." The study showed that official and civil society not only had disputes during this event but also formed two groups with different characteristic features. Notably, the government discourse network showed tighter intra-group connections, while civil society lacked sufficient coordination and cooperation both online and offline. The intermediary (Weibo) provided channels for opinion exchange, and through Susceptible-Infected (SI) simulation, demonstrated the intermediary's (Weibo) structural role in information diffusion in conversational networks.

2.12 Digital Libraries, Curation, and Preservation

Peter Organisciak from the University of Denver used algorithms to identify duplicate works at different granularity levels in the HathiTrust Digital Library, ranging from variant content of existing versions (same work, different expression) to variant presentations of identical content (same expression, different manifestation) to different printed copies (same manifestation, different item). Case analysis revealed that duplicate work adjustment in large digital libraries presents tremendous complexity and challenges. Ayse Gursoy from the University of Texas at Austin analyzed over 200 patch note documents for the League of Legends digital game from 2009 to the present, examining various character changes recorded during the game's development and how these changes evolved over time and their impacts. Case analysis results showed that in dynamic, com-

plex digital objects, changes can have different functions. Understanding how complex digital objects evolve over time and how these objects are constructed as mutable objects can provide insights for developing more nuanced models. Benedict Salazar Olgedo from the University of California, Irvine proposed a Bourdieusian approach to understanding video game preservation as an emergent field, whose discourse and practices are shaped by ontological differences and conflicting power structures. This can be illustrated through a series of evaluations that view video games on one hand as permanent like material objects, and on the other hand as ephemeral. Video game preservation initiatives may tend toward one of these extremes, focusing on ensuring playability or documenting gameplay. By classifying participants, we can better understand these competing values, thereby revealing power relations within them. The author concluded that the future of video game preservation depends on whether the next generation of video game archivists can form communities with clear shared values and philosophies.

2.13 Social Media Text Mining and Sentiment Analysis

Brooke Auxier from the University of Maryland studied 92 individual influencer accounts on Twitter with over 10K followers, using the VADER sentiment analysis package and psycholinguistic text analysis tool LIWC to analyze emotional distribution patterns and language use styles in health influencers' tweets. The study found that promotional themes were most common in health-related topics, followed by health/exercise information, motivation, self-disclosure, and sports-related content. In terms of emotional usage, no statistically significant differences were found in negative emotions between male and female influencers' tweets, but female influencers' tweets contained more positive emotions than male influencers' tweets. No statistically significant differences were found in positive or negative emotion usage between general fitness accounts and those self-identified as fitness "trainers" or "coaches." Sijing Chen from Wuhan University used the Hurricane Harvey Twitter dataset provided by the University of North Texas Library to explore interactions among social media users' emotions, location, and time during disasters and their impact on tweet retweeting patterns using the SentiStrength algorithm. Radhika Garg from Syracuse University analyzed emotional distributions of posts and their replies on social platforms to explore how discussions on Reddit influence users' decisions to use or abandon wearable devices. Based on 6,680 posts and 50,867 corresponding comments posted on the Reddit subreddit r/androidwear between December 2015 and December 2017, sentiment analysis found that post emotions significantly influenced comment emotions. Most comments on positive and neutral posts were positive, and even when many negative comments appeared in response to negative posts, these posts still saw fact-driven neutral discussions. Additionally, post and comment emotions influence users' decisions to use or abandon devices—when other Reddit users provide solutions or discussions receive mostly positive or fact-driven neutral comments, people decide to continue using their devices. Amir Karami from the University of Southern California used senti-

ment analysis and topic modeling techniques combined with qualitative coding to propose a popularity analysis framework for discovering reasons for politicians' popularity. He collected and analyzed 4.5 million tweets related to U.S. politician Senator Bernie Sanders on Twitter, using LIWC to identify positive tweets and LDA modeling for topic analysis of positive tweets, obtaining topic distributions across 10 economic dimensions and identifying eight economic reasons for Sanders' popularity: college tuition, work income, benefits, employment and income, minimum wage, monetary policy and the Federal Reserve, rights, and immigration.

2.14 Data and Information in the Public Sphere

Yong Ming Kow from City University of Hong Kong conducted semi-structured interviews (task-based information search and in-depth interviews) with 21 Hong Kong residents recruited through snowball sampling (12 male, 9 female, aged 20-62) to understand residents' preferred sources and search paths for news information. Using grounded theory to code interview data followed by iterative inductive analysis, they explored how residents handle false and erroneous information in news. Results showed that most citizens are aware of misinformation but only a small portion seek additional sources to cross-compare with erroneous information. There is self-rationality in news reading, with residents preferring brief news from authoritative sources. The study recommended that interaction designers should provide more rich news content to help eliminate misinformation and consider long-term solutions such as strengthening citizens' news literacy. Kalpana Shankar from University College Dublin reviewed literature on public-private partnerships in genealogical data supply services in recent years to explore challenges and development trends of public-private partnerships in data services. The study found that public-private partnerships in genealogical data services arise because libraries and archives hold large amounts of genealogical materials of interest to private enterprises and historical researchers but lack the ability to widely disseminate resources. Tensions and access disparities exist in public-private partnerships for genealogical data services; service terms in collaborations include exclusive and confidentiality agreements; and data services require extensive planning before partnerships, considering management, storage, access costs and benefits to realize data value. Di Wang from Wuhan University used online questionnaires to collect Chinese citizens' sociodemographic information, awareness of government open data, and demand themes for government open data, quantifying demand levels for government open data using Likert scales. Through data analysis and testing, three hypotheses were proven (user, data type, and utilization rate influence each other), aiming to provide governments with relationships among open data, users, and usage to improve services. Results showed that currently few people understand and use government open data portals; users with different sociodemographic characteristics show varying degrees of demand for different topics in open data; and daily life information and anti-corruption are two main types of citizen utilization of government open data. The study recommended that governments need to

increase citizen awareness of open data websites and improve dataset quality to stimulate citizen usage intentions.

2.15 Engaging with Multimedia Content

Toine Bogers from Aalborg University Copenhagen scraped 2,266 posts about video games from three Reddit groups (/r/gamingsuggestions, /r/gamesuggestions, /r/tipofmyjoystick) and identified 521 posts expressing video game information needs as research data. Using open coding and axial coding, they summarized seven primary elements and 41 secondary elements of users' video game information needs expression. Statistical analysis found that content, metadata, experience, context, and interactivity were five elements frequently mentioned in game information searches, with content and metadata being descriptive elements frequently used when searching for games with known content features, while experience, context, and interactivity were descriptive elements frequently used when searching for similar games. Kayla M. Booth from the University of Pittsburgh selected sports and nutrition as health topics, conducting semi-structured interviews with 30 U.S. adolescents aged 13-18 to explore how they evaluate high-risk video-based health information on social media from two aspects: external cues like platform and information source, and visual and content interaction factors based on the videos themselves. The study found that among external cues, platform advertisements and comments on videos, especially negative and exaggerated comments, negatively impact adolescents' evaluation of video-based health information, while detailed and interesting video content receives higher ratings among adolescents. Video image quality and quantitative metrics like likes and views also influence adolescents' evaluation of video-based health information on social media. During viewing, the competence and professional level of people appearing in videos similarly impact adolescents' evaluations. Sally Jo Cunningham from the University of Waikato analyzed interview data from all music behavior studies at the university between 2002-2016 to explore whether streaming services have transformed personal music collections. The study found that personal music collections mainly divide into three phases: 2002-2004 mixed physical and digital collection phase, 2006-2007 transition to digital collection phase, and up to 2016 large-scale music collection and music streaming phase. The third phase primarily involves digital music collection, where boundaries between users are very blurred compared to previous phases, and users find it difficult to assess collection size. This phase also facilitates convenient music collection anytime without users worrying about transfer costs. The summary finding is that current streaming services have indeed transformed how users collect music.

2.16 Understanding Online Behaviors and Experiences

Tingting Jiang from Wuhan University analyzed 3,999,020,0 user clickstream data from server log files of a website over two months, combined with analy-

sis of news novelty and popularity scraped from the website's homepage. Using the Mann-Whitney U test to examine relationships between news novelty/popularity and user click behavior, they investigated how news timeliness and popularity affect user click behavior. Analysis of click data found that both indicators influence news headline selection, and a "herding effect" shows that when many people believe information, individuals tend to believe it. These results have important implications for news providers in creating effective headlines and publishing and disseminating news more responsibly. Dan Wu from Wuhan University selected the "Digital Dunhuang" platform as a case study, using two data collection methods: eye-tracking data recorded by Tobii Studio, collecting eye movement experimental data from 40 participants (20 male, 20 female, from 19 schools and 32 majors, with education levels ranging from freshman to graduate student). Eye movement features focused on fixation points, saccades, and derived metrics, with experimental tasks including free browsing, visual search, and interaction. The second method was a questionnaire survey based on the System Usability Scale (SUS), with corresponding questionnaires for each task. Analysis of 31 subjects with gaze sampling rates above 85% aimed to combine users' gaze characteristics with experience characteristics to build an evaluation model for testing digital humanities platform usability and identify relationships between them. Results showed that more fixations and longer total fixation time indicate higher user interest in AOIs, suggesting higher attention levels; fewer saccades, shorter total saccade distance, and shorter saccade time intervals indicate fewer visual searches, easier target finding, and user satisfaction with platform structure and layout, indicating high efficiency. User satisfaction had no significant effect on average fixation time or average saccade distance.

2.17 Algorithms at Work

Jiawei Chen from the University of Pennsylvania used experimental and questionnaire methods to conduct a technical tracking survey of 23 U.S. college students (installing apps on participants' phones to collect data). The study examined users' behaviors, attitudes, and needs when using ride-sharing applications in different context-aware scenarios (requesting rides or driving to help others), aiming to explore factors influencing community co-production to improve recommendation algorithms. Using clustering and simulation analysis, the study found that beyond traditional factors like time, location, and interpersonal relationships, community parameters and spatiotemporal contexts such as user resources, emotions, abilities, and vehicle size are important factors influencing community co-production. Results indicate that when recommendation algorithms lack contextual parameters, community parameters and spatiotemporal background factors can serve as alternative supplements. Mohammad Hossein Jarrahi from the University of North Carolina at Chapel Hill conducted an empirical study through video interviews, policy document analysis, and inductive analysis of 26 employee profiles and 98 forum topics on the Upwork freelance platform to explore whether algorithmic management (automatically

assigning different resource tasks to platform workers) creates information asymmetry. The study aimed to understand how employees on online digital labor platforms comprehend and utilize platform algorithms. Results showed that platform algorithmic rules affect online workers' rating scores, search rankings, and salary quotes, with information asymmetry impacting workers' control over work. However, online workers are not passive recipients of algorithms; those with algorithmic management capabilities can perceive, circumvent, and even manipulate algorithms. Algorithmic management is a sociotechnical process resulting from continuous interaction between platforms and users; literacy in understanding and utilizing algorithms can be seen as core to individual competitive advantage in the global market. Alan Rubel from the University of Wisconsin-Madison analyzed two cases—the "2017 Facebook anti-Semitic automatic ad targeting incident" and the "Wisconsin Supreme Court Justice using COMPAS proprietary risk assessment tool to write criminal suspect investigation reports"—proposing the concept of "institutional laundering" as a metaphor for how organizational institutions using algorithmic decision-making systems transfer all consequential responsibilities to the algorithm itself to evade ethical constraints and obscure their own responsibilities. The study explained the relationship between "institutional laundering," "cover-up behavior," and "responsibility absence," showing that using algorithmic systems for important decisions poses potential ethical risks.

2.18 Innovation and Professionalization in Technology Communities

Yao Cai from Nanjing University of Science and Technology used 16,732 articles published in 86 intelligence and library science journals indexed in Web of Science between 2006-2015 as a dataset, conducting co-word analysis of keywords to reveal iSchool research topics and their evolution. The study aimed to explore whether iSchool development remains centered on "information, technology, and people." Nine interdisciplinary research themes were identified: "library," "trust," "knowledge management," "social network," "bibliometrics," "information retrieval," "information technology," "text mining," and "information literacy." Results showed that research on information has always been the main line of iSchool development, technology is an important factor promoting iSchool development, and providing better services for users is the ultimate goal of iSchool development. Guo Freeman from Clemson University conducted 12 semi-structured in-depth interviews averaging 80 minutes each with amateur independent game developers to obtain information about how independent game developers innovate, with iterative coding of interview content. The study aimed to explore an informal collective learning network composed of distributed individuals through research on independent game developers' innovation patterns, constructing an innovation ecology that enables bottom-up participation in creative technology practices and revealing the operational mechanisms of innovation ecology. Results showed that solving problems in various fields motivates developers to innovate; help, communication, and encouragement from communities support innovation; and political and policy fac-

tors can promote or hinder innovation. Christine T. Wolf from IBM Research conducted semi-structured interviews with 11 machine learning developers at a large global technology company to investigate their daily work practices and products used at work. Using inductive and thematic analysis techniques, the study explored developers' interest development process in machine learning, barriers encountered when entering the field, and information patterns used in these processes. The study aimed to explore the role of information in career choice and professional identity and the relationship between information use and professional identity. Results showed that different types of information stimulate individuals' interest in machine learning; information obtained from colleague communication and online platforms at work helps developers solve work problems; and information use serves as an organizing principle that can connect a professional group.

2.19 Information Behaviors on Twitter

Asel Addawood from the University of Illinois collected 146,403 relevant tweets on two health topics (MMR vaccines and healthy diet) from Twitter between January 1, 2016, and November 28, 2016. Using a mutually exclusive domain dictionary, URLs were categorized into 12 categories: news, video, fake news, social media, blog, business, government, science, educational institution, health magazine, national or state professional medical association, and health insurance. Influential Twitter users were identified at individual and organizational levels based on Klout scores (top 25% considered influential) and divided into different categories. By analyzing URL categories, user categories, and their relationships in relevant user messages, the study explored relationships between different user categories and their URL sharing patterns. Results showed that different types of influential users participate in different health topics, and influence is not obtained by sharing scientific or factual evidence but can be obtained through personal relationships. For example, when influential users share fake news (such as users discussing vaccine-related topics sharing fake news URLs being more than twice as likely as those discussing healthy diet). Zheng Gao from Indiana University Bloomington used movies produced before October 2016 from IMDb as a data source, using IMDb user average ratings as an indicator of critical performance and return on investment as financial performance. The study quantitatively analyzed basic features of movie metadata and comprehensively calculated complex movie features, using Support Vector Machine (SVM) to predict movie success based on identified features. Results explained three common phenomena in the film industry: family dramas are more likely to attract audiences across all genres; a movie's success largely depends on the past career success of its actors; and stable collaboration between directors and actors is more likely to achieve long-term movie success, especially in series. Yingya Li and Bei Yu from Syracuse University used conclusion section finding sentences in biomedical abstracts as research objects, randomly selecting 1,100 biomedical abstracts containing five common health topics from PubMed for randomized clinical trial research design grouping (1,000 as training set, 100 as

test set). Comparing rule-based methods and machine learning methods across metrics including precision, recall, and macro-averaged F1 score, experimental results showed that both rule-based methods and bag-of-words machine learning methods achieve high precision, with simple rule-based methods performing better. This indicates that although advanced machine learning methods can capture main patterns, human experts may still perform better on such specialized tasks. Lu Xiao and Xin Huo from Syracuse University used an open-access corpus provided by Rutgers University's data mining group, selecting 271 text excerpts containing reasons (1,124 sentences total) and annotating text with authoritative claims. The study examined whether correlations exist between reasons and authoritative claims in online disseminated content. Results showed only 42 sentences contained authoritative claims (3.7% of total), indicating that when users present reasons to others in these activities, they rarely attempt to enhance their credibility.

2.20 Data Mining and Natural Language Processing

Zheng Gao from Indiana University Bloomington used movies produced before October 2016 from IMDb as a data source, using IMDb user average ratings as an indicator of critical performance and return on investment as financial performance. The study quantitatively analyzed basic features of movie metadata and comprehensively calculated complex movie features, using Support Vector Machine (SVM) to predict movie success based on identified features. Results explained three common phenomena in the film industry: family dramas are more likely to attract audiences across all genres; a movie's success largely depends on the past career success of its actors; and stable collaboration between directors and actors is more likely to achieve long-term movie success, especially in series. Yingya Li and Bei Yu from Syracuse University used conclusion section finding sentences in biomedical abstracts as research objects, randomly selecting 1,100 biomedical abstracts containing five common health topics from PubMed for randomized clinical trial research design grouping (1,000 as training set, 100 as test set). Comparing rule-based methods and machine learning methods across metrics including precision, recall, and macro-averaged F1 score, experimental results showed that both rule-based methods and bag-of-words machine learning methods achieve high precision, with simple rule-based methods performing better. This indicates that although advanced machine learning methods can capture main patterns, human experts may still perform better on such specialized tasks. Lu Xiao and Xin Huo from Syracuse University used an open-access corpus provided by Rutgers University's data mining group, selecting 271 text excerpts containing reasons (1,124 sentences total) and annotating text with authoritative claims. The study examined whether correlations exist between reasons and authoritative claims in online disseminated content. Results showed only 42 sentences contained authoritative claims (3.7% of total), indicating that when users present reasons to others in these activities, they rarely attempt to enhance their credibility.

2.21 Informing Technology Design Through Offline Experiences

Julia Cope from the University of Pittsburgh conducted semi-structured interviews with 12 firefighters from California, Pennsylvania, and Maryland to obtain data about their navigation training processes, navigation strategies used in rescue missions, problems encountered, and solutions. Using grounded theory to code interview information, the study explored how firefighters make judgments about surrounding information and develop spatial search and rescue operation strategies during emergency rescues, incorporating results into the design of personalized navigation assistance tools for firefighters. Results showed that firefighters primarily use various navigation strategies for four purposes: quickly establishing search paths in rescues, improving visual visibility and range, creating cognitive maps through deduction of building internal structures, and making orientation (direction determination) decisions during rescues. Dede Ma and Pengyi Zhang from Peking University used online questionnaires to survey 1,129 ride-sharing application users in four Chinese cities, examining passengers' emotional experiences (happiness, satisfaction, surprise, anxiety, disappointment, anger) when using ride-sharing applications and the relationships among emotional experiences, user characteristics (gender, age, education, usage frequency/week, expenditure/week), usage contexts (time, comfort, navigation accuracy, road conditions, urgency, presence of other passengers), and application interaction satisfaction. The study aimed to provide preliminary understanding of passengers' emotional experiences to improve such applications. Results showed that passengers experience more positive than negative emotions when using ride-sharing mobile applications; users with different negative emotional experiences have different emotional venting paths and action tendencies; usage context, interaction satisfaction, and user characteristics relate to passengers' emotional experiences—when interactions with ride-sharing applications are more successful, passengers have more positive emotions, while female and frequent users have more negative emotions. Swathi Jagannath from the University of the Sciences observed and video-recorded 44 instances of electronic flowchart documentation during medical resuscitation at a pediatric teaching hospital in the mid-Atlantic region of the U.S., including challenges encountered and solutions, and interviewed 24 nurse documenters. The study aimed to understand nurse documenters' information behaviors and attitudes toward newly implemented electronic flowcharts during medical resuscitation, exploring differences in information behaviors and attitudes during the transition from paper to electronic documentation in high-risk medical environments to improve electronic flowcharts and support design guidelines for their use during resuscitation. Results showed that electronic documentation in complex medical environments has advantages in supporting continuous patient care but faces challenges in flowchart navigation, automatic filling of resuscitation data, and real-time documentation creation practices.

2.22 Digital Tools for Health Management

Elizabeth V. Eike from the University of California assessed eating disorder (EDs) severity among 24 female college students using four questionnaires including EDE-Q6.0 and interviewed them about whether they would recommend general diet and health apps to other patients for recovery and whether they would use specialized EDs treatment apps. The study found that most users use general diet and health apps but would not recommend them to other patients, and most would not use specialized EDs treatment apps. Through coding interview data, the study found that not knowing about their existence is the main reason users don't use specialized EDs treatment apps, along with reasons including being unpopular and unfamiliar, unwillingness to use them, and poor usability of these apps. Jomara Sandbulte from Pennsylvania State University conducted focus group interviews with 21 adults and 16 older adults to explore and research factors that promote sustained health information sharing between different generations in families. The study found that close relationships with relatives naturally lead to sustained health information sharing; when in good health, sustained health information sharing may sometimes feel troublesome to family members, but when physical discomfort occurs, such sharing becomes particularly important. The most prominent finding of this study is that turning points motivate and promote sustained health information sharing among family members. Turning points refer to sudden major health events among family members, such as sudden serious illnesses. Additionally, family genetic history and open family environments also promote sustained health information sharing among family members. Shikun Zhang from Carnegie Mellon University recruited 200 research participants aged 18 and older residing in the U.S. on the MTurk platform to study users' sleep behaviors and their impact on sleep, as well as strategies and tools users adopt to improve sleep quality through three-part surveys including demographic surveys, user sleep quality surveys, and user strategy surveys. Qualitative analysis of open-ended questions in the questionnaires revealed users' sleep behaviors and improvement tools and strategies. Quantitative analysis of structured questions in the questionnaires found that demographic characteristics such as age, occupation, and sleep schedules, as well as related medications, may affect sleep quality. Current sleep improvement technologies show no significant effect on improving users' sleep quality.

2.23 Environmental and Visual Literacy

Jacob Abbott from Indiana University used online surveys and semi-structured interviews to survey 31 participants from different groups (environmental science PhD students, architecture company owners, student sustainable development committee meeting participants, undergraduate students) about recycling practices, resource conservation habits, technology use, transportation methods, environmental literacy, and attitudes toward environmental issues. The study aimed to create a mobile carbon footprint calculator prototype based on this data. Results showed that most participants have needs to measure

resource consumption and its environmental impacts, believing measurement results should be easy to read. Based on this, they built the carbon footprint calculator prototype CreEn, which uses user input and real-time data from IoT devices to measure energy consumption and output personalized carbon footprint reports. Monica Maceli from Pratt Institute conducted focus group interviews with four archival practitioners, including voting on predefined data and dashboard function lists and drawing visual appearances of dashboards in pairs, to explore archival practitioners' needs and challenges for data dashboards in environmental monitoring and propose a dashboard model based on user needs and challenge data. The study showed that archival practitioners face technical, resource, physical space, data background knowledge, communication, and organization-related challenges when using data dashboards, with technology and resources being dominant. Yong Ying Joanne Tan and Tin Seong Kam from Singapore Management University used statistical testing methods to calculate local indicators of spatial association (LISA) for Singapore public housing postal codes and analyzed them in R environment, visualizing by housing type (1-2 room apartments, 3-room apartments, 4-room apartments, 5-room and executive apartments) to explore LISA applications in electricity consumption data, identifying electricity consumption anomalies and spatial distribution of high-consumption households and local electricity consumption pattern clusters.

2.24 Addressing Social Problems

Yuanye Ma from the University of North Carolina at Chapel Hill used China's proposed social credit system as an example to explore privacy issues from a cultural perspective, arguing that privacy is a cultural concept whose meaning in Chinese culture has not been fully explored, and proposing the importance of comparing privacy cultures across countries. This conceptual exploration opens new research areas for future privacy work. Sally Sanger from the University of Sheffield used semi-structured interviews to survey 25 users from different backgrounds about their information seeking behaviors in online alcohol support group forums, coding data with NVivo 11 and developing themes using Braun and Clarke's thematic analysis method. Results showed that the process of using online alcohol support group forums to form ideas about drinking problems matches Bates' "berrypicking" information search model, with information seeking behaviors changing over time and users using multiple sources and search techniques to refine ideas.

---

3. Summary

3.1 Research Scenarios

The transformation and development of data science and data-driven research paradigms have prompted integration between library and information disciplines and other disciplines, discovering research questions in new research sce-

narios. Research scenarios covered in this conference include bioscience, astronomy, earth science, software engineering, smart personal assistants, smart homes, digital games, medical health, mobile payments, digital music, digital humanities, transportation, environmental protection, firefighting, ride-sharing, and other scenarios, with research focusing on user, technology, and information-related issues in these contexts.

3.2 Research Methods

Main research methods used in conference papers include questionnaire surveys, semi-structured interviews, focus groups, experimental observation, grounded theory, thematic analysis, statistical analysis, Photovoice, social network analysis, case studies, machine learning, data mining, meta-analysis, eye-tracking experiments, and other methods, with grounded theory, semi-structured interviews, and questionnaire surveys being the most predominant methods.

3.3 Research Data

Over 80% of papers in this conference collected data through various methods, discovering features and patterns behind data through coding, statistical analysis, or text analysis methods. Therefore, research data acquisition is an extremely important process in these studies. Research data used in conference papers includes both data scraped from various online applications, such as social media data (e.g., Twitter, Weibo), online discussion community data (e.g., Reddit), academic paper data (WoS), Internet Movie Database (IMDb) data, user review data (e.g., Amazon, Bestbuy), product introductions, and user manuals, as well as data obtained by researchers through structured interviews and questionnaire surveys.

References

[1] TAYLOR N G, CHRISTIAN-LAMB C, MARTIN M H, et al. Information in contemporary society, 14th International conference, iConference 2019, Washington, DC, USA, March 31-April 3, 2019, proceedings[C/OL]. Washington: Springer, 2019[2019-05-08]. https://doi.org/10.1007/978-3-030-15742-5_1.
[2] IDEALS. iConference 2019[EB/OL]. [2019-05-08]. https://www.ideals.illinois.edu/handle/2142/102118.
[3] NRC awards and honors dataset[EB/OL]. [2019-05-08]. http://sites.nationalacademies.org/pga/resdoc/pga_

Author Contributions

Gong Hongcun: Data collection and organization, paper writing
Shen Yutian: Data collection and organization, paper writing
Cao Gaohui: Paper topic selection, conceptual design, paper writing and revision

English Title and Abstract

"Inform, Include, Inspire"—A Summary of iConference 2019

Abstract: [Purpose/significance] iConference is an important international conference in the field of information science. By combining and summarizing the accepted papers in iConference 2019, this paper aims to provide reference for future related research in information science. [Method/process] From the four dimensions of research method, research topic, research process and research finding, this paper summarized reviewed 77 accepted papers on 24 themes in iConference 2019. [Result/conclusion] As the annual conference of iSchool, iConference's topic covers the latest achievements of theoretical research and application in the field of information science. From the accepted papers, we find that the data-driven research paradigm provides an opportunity for the integration and development of information science and other disciplines, and promotes libraries and information disciplines to discover research problems in new research scenarios; semi-structured interviews, questionnaires, grounded theory, statistical analysis and other qualitative and quantitative methods are the main research methods; research data acquisition is an extremely important research process.

Keywords: iConference; conference review; data-driven; information science

Note: Figure translations are in progress. See original paper for figures.