

Chapter 23

Automated Data Extraction from Online Social Network Profiles: Unique Ethical Challenges for Researchers

Sophia Alim
Barnardos, Bradford, UK

ABSTRACT

As the use of online social networking (OSN) sites is increasing, data extraction from OSN profiles is providing researchers with a rich source of data. Data extraction is divided into non-automated and automated approaches. However, researchers face a variety of ethical challenges especially using automated data extraction approaches. In social networking, there has been a lack of research that looks into the unique ethical challenges of using automated data extraction compared to non-automated extraction. This article explores the history of social research ethics and the unique ethical challenges associated with using automated data extraction, as well as how these impact the researcher. The author's review has highlighted that researchers face challenges when designing an experiment involving automated extraction from OSN profiles due to issues such as extraction methods, the speed at which the field of social media is moving and a lack of information on how to deal with ethical challenges.

INTRODUCTION

Online social networks (OSNs) have become a popular way for users to interact with their friends via the display of a user's personal details, friendship connections and interactions with other users on OSN profiles. In December 2012, Facebook, the leading OSN, had more than one billion active users (Facebook, 2013).

Boyd and Ellison (2008) have described three ingredients that constitute an OSN:

1. Allowing a user to make a public or semi-public profile inside a system that is bounded. A bounded system represents a set of activities that are interrelated and come together to make a single entity, which in this case is the OSN;

DOI: 10.4018/978-1-4666-6433-3.ch023

2. Bringing together a list of other users with whom they share a connection and allowing the user to view these;
3. Allowing users to travel along their set of connections and the connections made by other users within the system.

Depending on the users' attitudes on privacy, data on a user's OSN profile, which includes personal details, friendship connections and interactions with other users, can provide a rich source of useful information for extraction for a variety of researchers, from different domains that range from computer science to sociology.

Various approaches are available for extracting publically available OSN profile data. Extraction approaches can be split into two separate categories: non-automated and automated.

Non-automated approaches utilise surveys and interviews. They have been used in research studies such as the ones by Dwyer, Hiltz and Passerini (2007), Gibson (2007), Govani and Pashley (2005) and Strater and Richter (2007). Non-automated approaches include manual scrapping of OSN profile data, having conversations with people, joining forums to discuss issues of interest and listening in on conversations between other people.

During the last couple of years, more studies have been done with data extracted using automated extraction approaches, e.g. by Catanese, De Meo, Ferrara, Fiumara and Provetti (2012); Caverlee and Webb (2008); Pfeil, Arjan and Zaphiris (2009).

At present, there is significant debate as illustrated by BBC News (2011); Giles, Sun and Council (2010); Zimmer (2010) about the issue of ethics in the automated extraction of publically available OSN profile data by researchers. This is due to the desire of some OSN users for control over their personal details, their curiosity about who can access their profile data and need to know what their personal details are used for. In the past, researchers could use automated means to extract public OSN profile data without informing the profile user.

Previous research that was carried out in the area of ethics and accessing OSN profile data by Wilkinson and Thelwall (2011); Wilson, Gosling and Graham (2012); Zimmer (2010) has discussed the ethical issues of extracting publically available profile data, but did not explore the unique ethical challenges that automated extraction from OSNs brings in comparison to non-automated extraction.

Wilson et al. (2012) explored automated extraction via data crawling and ethics in their overview of Facebook, whilst Wilkinson and Thelwall (2011) described techniques to access personal information including automated data extraction via specialist software. The Wilkinson and Thelwall (2011) study surveyed areas such as anonymity and informed consent with regards to extracting publically available OSN data and using it for research in an ethical manner. However, the study did not focus on automated data extraction or explore issues such as the storage of extracted data, OSN policies and the position of OSNs with regards to automated data extraction and the distribution of datasets. Zimmer (2010) explored the mistakes made regarding ethics when the *Tastes, Ties and Time* research study took place using non-automated approaches to extract OSN profile data.

This paper looks into the unique ethical challenges of using automated approaches for the extraction of OSN profile data and how these impact researchers.

The rest of the paper is organised as follows: Firstly the paper provides an overview of the history behind social research ethics and explores the motivation for ethical guidelines to be present in human research. The scene is set for an exploration of ethical issues associated with non-automated and automated methods of data extraction from OSN profiles and highlights the unique ethical challenges in using automated methods. This leads to a discussion of the challenges researchers have to face with regards to ethics. Finally the paper is concluded and ideas for future research are presented.

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/automated-data-extraction-from-online-social-network-profiles/117040

Related Content

Vengeance Culture Online: A Qualitative Meta-Analysis and Position Paper

Shalin Hai-Jew (2014). *Human Rights and the Impact of ICT in the Public Sphere: Participation, Democracy, and Political Autonomy* (pp. 64-93).

www.irma-international.org/chapter/vengeance-culture-online/112167

Tools for the Diagnosis of Corruption Symptoms in Government Entities: The Colombian Case

Diego Alejandro Peralta Borray, Nelson Andrés Molina Roa, Monica Consuelo Rodriguezand Sebastian Zapata (2023). *Management Strategies and Tools for Addressing Corruption in Public and Private Organizations* (pp. 171-196).

www.irma-international.org/chapter/tools-for-the-diagnosis-of-corruption-symptoms-in-government-entities/325051

Executive Agency “General Labour Inspectorate“: General Characteristics and Structure

Galina Yolova (2023). *Legal and Economic Aspects of State Control Over Compliance With Labor Legislation* (pp. 44-56).

www.irma-international.org/chapter/executive-agency-general-labour-inspectorate/328424

Open Data for Open Justice in Seven Latin American Countries

Sandra Elenaand François van Schalkwyk (2017). *Achieving Open Justice through Citizen Participation and Transparency* (pp. 210-231).

www.irma-international.org/chapter/open-data-for-open-justice-in-seven-latin-american-countries/162844

Affirmative Action in the United States and Positive Action in the European Union in the Context of Comparative Law

Anna-Maria Konsta (2015). *Human Rights and Ethics: Concepts, Methodologies, Tools, and Applications* (pp. 526-554).

www.irma-international.org/chapter/affirmative-action-in-the-united-states-and-positive-action-in-the-european-union-in-the-context-of-comparative-law/117047