



Ethics and Legal Implication of Web Scraping: An Analysis and Tool for Developing Risk Assessment Model

Agenda

Project Overview

Business Case

Statement of the Problem

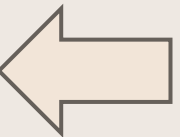
Methodology

SOP vs Findings vs Conclusions

Recommendation

PROJECT OVERVIEW

Web scraping, the automated extraction of data from websites, has become a widespread practice with profound implications for ethics and legality. This research aims to delve into the ethical and legal aspect surrounding web scraping, exploring challenges, awareness levels, and proposing a model for responsible and compliant practices.



Business Case

1

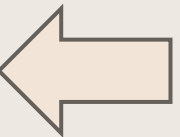
Rapid Growth of Web Scraping

2

Lack of Standardized Guidelines

3

Privacy Concerns



Statement of the problem

1. What are the most significant challenges that emerge due to the lack of established ethical guidelines and unclear legal regulations regarding web scraping?
2. What is the level of the awareness of the respondents about ethical and legal aspects of web scraping in terms of:
 - 2.1. Attitude
 - 2.2. Social Pressure
 - 2.3. Self-Efficacy



Statement of the problem

3. What is the level of agreement of the respondents on the Ethical and legal Web aspects of web scraping in terms of:

3.1. Basic Rights

3.2. Common Good Impact

3.3. Legal Compliance

4. Is there a significant relationship between the level of awareness and level of agreement on ethical and legal aspects of web scraping for respondents?



Statement of the problem

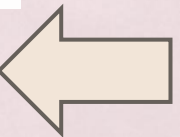
5. Which web scraping tool demonstrates the highest alignment with ethical and legal considerations?
6. What model can be developed that can be solved the ethical and legal implications of web scraping.





RESEARCH DESIGN

Quantitative Design, researchers found it compatible to know the challenges pose by the lack of ethical and legal guidelines, the level of awareness, level of agreement and the significance of the model proposed in understanding the ethical and legal web scraping for the students.



Statement of the Problem	Summary of Findings	Conclusions
<p>1. What are the most significant challenges that emerge due to the lack of established ethical guidelines and unclear legal regulations regarding web scraping?</p>	<p>Privacy concerns are a major challenge in web scraping, with 26 respondents identifying it as critical. This issue underscores the difficulty of protecting individual privacy while collecting large amounts of personal data. Jayan (2024) points out the complexities around informed consent, as users often accept terms without fully understanding them, questioning the validity of their consent.</p>	<p>Respondents are recognizing the challenge of setting ethical boundaries for using information from the web. This awareness marks a positive shift towards responsible data collection, emphasizing the need for ethical practices. This growing awareness could lead to the development of ethical guidelines and best practices for web-based data gathering.</p>



Statement of the Problem	Summary of Findings	Conclusions
<p>2. What is the level of the awareness of the respondents about ethical and legal aspects of web scraping in terms of:</p> <p>2.1. Attitude</p> <p>2.2. Social Pressure</p> <p>2.3. Self-Efficacy</p>	<p>Respondents have a strong understanding of the factors influencing their decision-making, with high scores in attitude, social pressure, and self-efficacy, reflecting positive beliefs and confidence. This aligns with Husband (2020) on the importance of ethical awareness and the balance of power in research, emphasizing that awareness shapes attitudes and actions.</p>	<p>Respondents show strong understanding of decision-making influences, with high scores in attitude, social pressure, and self-efficacy. They view the topic positively, feel societal and peer pressure, and are confident in acting effectively. This underscores the importance of understanding researcher-participant dynamics, especially regarding ethical power imbalances, emphasizing how awareness shapes attitudes and behaviors in decision-making.</p>



Statement of the Problem	Summary of Findings	Conclusions
<p>3. What is the level of agreement of the respondents on the Ethical and legal Web aspects of web scraping in terms of:</p> <p>3.1. Basic Rights</p> <p>3.2. Common Good Impact</p> <p>3.3. Legal Compliance</p>	<p>Respondents widely agree on the importance of basic rights, societal impact, and legal compliance in ethical decisions, recognizing the need to respect individual rights while following legal standards. This consensus supports Sobel (2021) on the role of ethics and legal adherence in data practices, showing a shared understanding of ethical implications.</p>	<p>Respondents largely agree on the ethical and legal considerations of web scraping in academics, emphasizing privacy, the common good, and legal compliance. There's a strong commitment to institutional guidelines and educational initiatives for ethical web scraping, prioritizing ethics over convenience.</p>



Statement of the Problem	Summary of Findings	Conclusions
<p>4. Is there a significant relationship between the level of awareness and level of agreement on ethical and legal aspects of web scraping for respondents?</p>	<p>A positive correlation exists between respondents' awareness and their support for ethical web scraping practices. This aligns with Krotov and Tennyson (2019), highlighting how awareness can promote responsible data collection practices, particularly in both academic and industry settings.</p>	<p>Awareness significantly influences attitudes towards ethical web scraping, with informed respondents more likely to support responsible practices. Education proves effective in raising awareness of risks, leading individuals to adopt ethical guidelines, make better decisions, and engage in responsible data collection.</p>



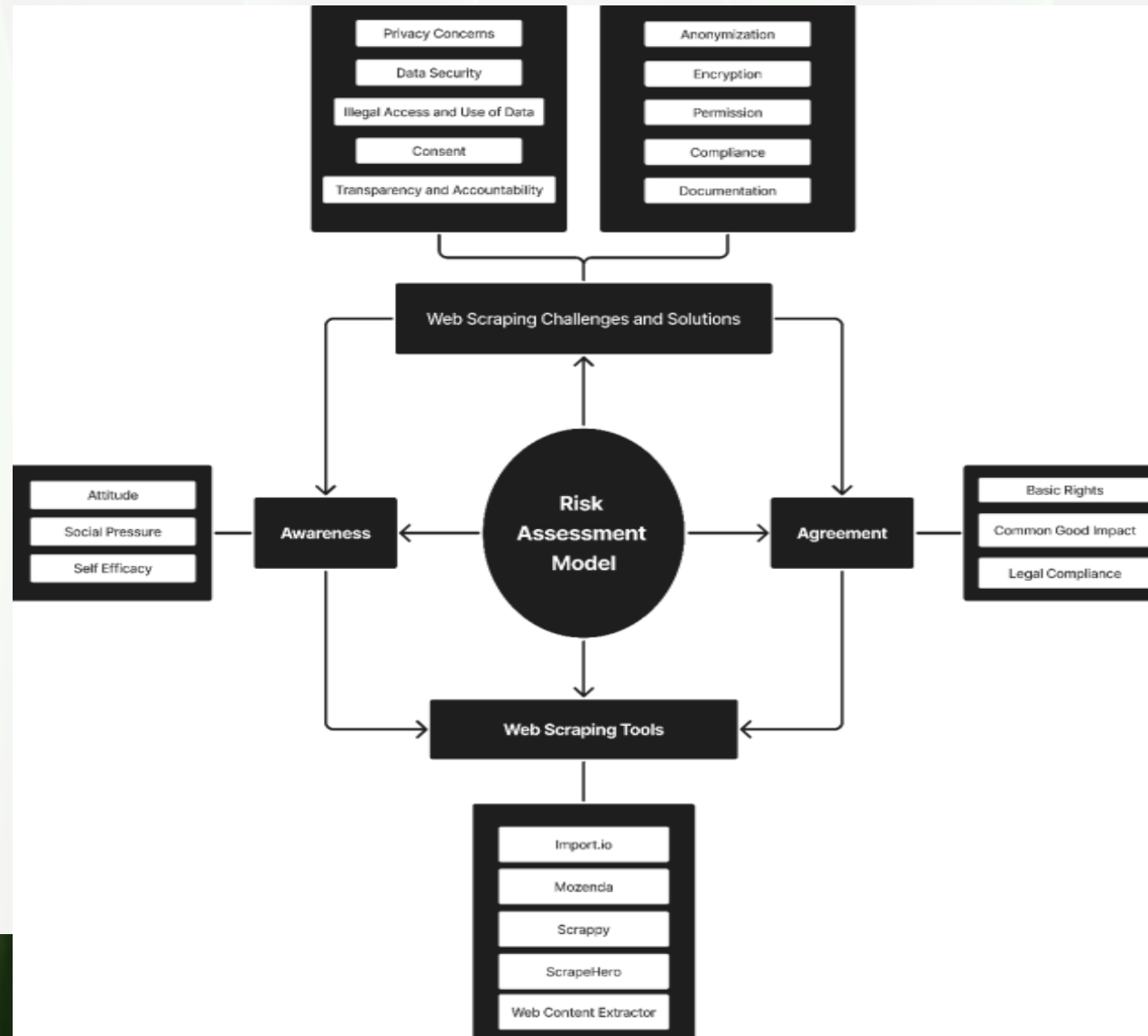
Statement of the Problem	Summary of Findings	Conclusions
<p>5. Which web scraping tool demonstrates the highest alignment with ethical and legal considerations?</p>	<p>Respondents favor specific web scraping tools like import.io, Mozenda, and ScrapeHero for their ethical and legal compliance, indicating a commitment to responsible data collection. This preference supports findings by Luscombe, Dick, and Walby (2021) on ethical considerations in web scraping, showing increased researcher awareness and a focus on integrity.</p>	<p>Researchers prefer web scraping tools like import.io and ScrapeHero due to their compliance with ethical and legal standards, reflecting a shift toward responsible data collection. Choosing such tools demonstrates a commitment to integrity and ethical standards in research practices.</p>



Statement of the Problem	Summary of Findings	Conclusions
6. What model can be developed that can be solved the ethical and legal implications of web scraping.	The Risk Assessment Model provides a guideline for ethical and legal web scraping, addressing data ownership, consent, and privacy. Rao (2022) highlights the model's role in navigating ethical complexities, offering structured guidance that upholds integrity in data collection and respects individual rights.	The Risk Assessment Model offers a clear framework for ethical and legal compliance in web scraping, addressing data ownership, privacy, and consent issues. It helps web scrapers act responsibly, balancing efficiency with ethics, and maintaining trust and accountability in research.

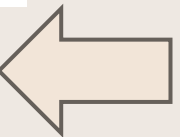


Risk Assessment Model



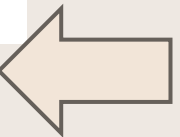
Recommendation

- Future web scrapers are encouraged to use the Risk Assessment Model to ensure alignment with ethical and legal standards, especially concerning privacy and data integrity. Staying informed on evolving laws and collaborating within the research community can help develop shared ethical norms.



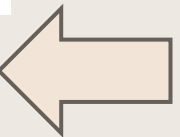
Recommendation

- Practitioners are advised to conduct regular ethical audits to ensure compliance with legal standards, focusing on transparency, consent, and data security. Clear ethical protocols can help web scrapers address potential issues and maintain accountability in data collection.



Recommendation

- Future researchers are encouraged to explore further dimensions of web scraping ethics, including its long-term effects on privacy and the influence of AI in ethical data collection. This will enhance understanding and contribute to the establishment of best practices in the field.



A close-up photograph of a green leaf, showing its veins and texture, positioned on the left side of the slide.

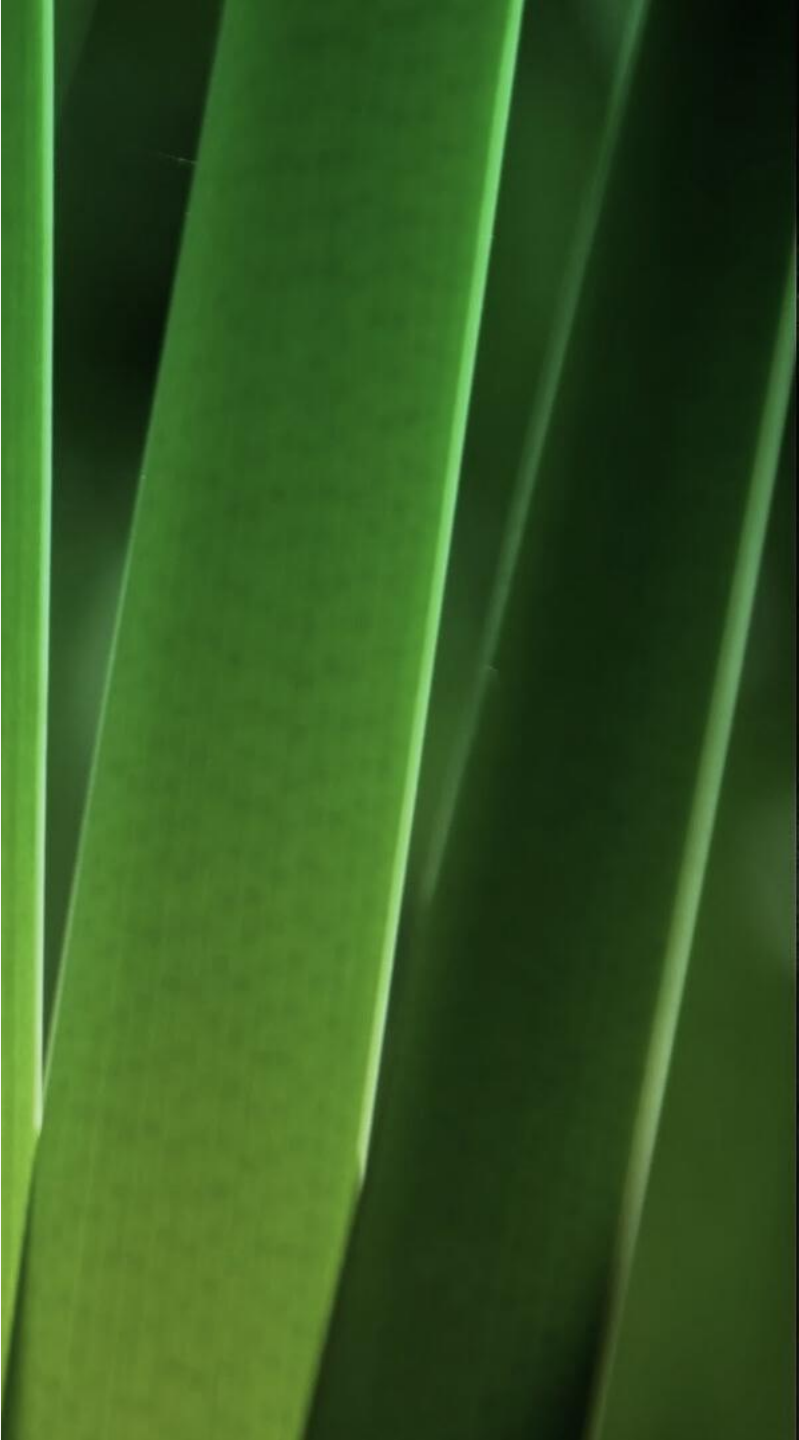
Thank you

BSIT 3-1

Cabral, John Loyd

Matudan, Levi

Mendoza, Kyle

A close-up photograph of a green leaf, showing its veins and texture, positioned on the right side of the slide.