Suggested Prototype Software to Extract Data from Social Media

Samaneh BORJI
Dr. Amir Reza Asnafi
Dr. Maryam Pakdaman Naeini
Introduction

This research focuses on social media data archiving.

What is social media archiving? Social media archiving is the process of capturing and storing social media content in a digital repository after it has been created.
Importance

Why is social media archiving important?

1. As a result of the introduction of social media, the nature of communication between humans has been transformed, with the medium establishing itself as the **primary means of communication** worldwide. At the same time, as making a **significant contribution** to data transmission, dissemination, and creation.
Importance

Why is social media archiving important?

2. Due to the **dynamic** and **unrestricted nature** of the Web, it is equally important and necessary to **preserve online data** in the same way that written documents are preserved.
Importance

Why is social media archiving important?

3. It is also possible to obtain vast amounts of information for research purposes by gaining access to archived data on social media.
Importance

Why is social media archiving important?

4. the **demise of a platform** may result in deleting user data from social media platforms. For example, Google announced the **termination of Google+** for customers in December 2018, and all material from consumer Google+ accounts would be removed. The deletion included not only **Google+ content** but also **comments published** on Google-owned services such as Blogger and YouTube.
Social Media archiving Software

1. The Internet contains different media forms, including photographs, videos, etc.
2. The amount of data available on social media is very high.

As a result, using powerful and specialized software to automate the data archiving process on social media is a prerequisite.
Problem Statement

1. According to the reports and the literature, social media archiving is still at a very early stage of development, with many countries attempting to establish pilot archive projects, despite the many difficulties they have faced so far. This process has also mounted a challenge to national archive organizations.

2. There is no a model including essential features of social media data archiving software.
Main Aim

The current study aims to analyze texts regarding extracting and archiving data from social media as a new, interactive communicative tool and examine similar software. Eventually, we proposed an optimal prototype of such software for local social media users based on the developed checklist. The main aim of this study is to present a prototype for extracting and archiving data from social media.
Other Aims of the Research

1. To identify the professional features of the current software used for extracting and archiving data from social media

2. To identify the necessity of the features of the current software used for extracting and archiving data from social media from experts’ point of view.
Research Questions

1. What are the professional features of the current software used for extracting and archiving data from social media?

2. What are the professional features of a practical prototype of software used for extracting and archiving data from social media?
Method

1. Literature Review
2. Reviewing Existing Software
3. Extracting the Features
4. Categorizing Features
5. Suggesting Prototype
6. Data analysis and statistical calculations
7. Gathering Expert opinions
Method

selected a group of 31 social media experts in information science

presented the software's features to the experts in a Likert scale format

performed CVR, exploratory factor analysis, and confirmatory factor analysis
### Concepts

**Archival Data**

### Professional features of the software

- Archiving different kinds of data formats
- Automatic 24/7 archiving
- Archiving with a visual link
- Archiving in a native format (with files, links, etc.)
- Archives classification
- Archive advanced search
- Restoring deleted data
- Sharing archives
- Automatic archive deletes within a specified amount of time
- Archived data in software for social media management
- Cloud Computing technology
- Creating an open portal for public access (public portal)
- Data extraction from other web sources (weblogs, websites, etc.)
## Result

<table>
<thead>
<tr>
<th>Concepts</th>
<th>Professional features of the software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software credit</td>
<td>Getting certificates from public and private institutions</td>
</tr>
<tr>
<td></td>
<td>Digital signature</td>
</tr>
<tr>
<td>Data Mining</td>
<td>Analyzing the extracted data</td>
</tr>
<tr>
<td></td>
<td>Presenting visual reports</td>
</tr>
<tr>
<td></td>
<td>Sharing the reports</td>
</tr>
<tr>
<td>Data security</td>
<td>Data coding (extracting, saving, sharing)</td>
</tr>
<tr>
<td></td>
<td>Assigning access levels</td>
</tr>
<tr>
<td>User management</td>
<td>Defining users and managers</td>
</tr>
<tr>
<td></td>
<td>Assigning roles to the users</td>
</tr>
<tr>
<td>Tariffs</td>
<td>Getting the complete version after paying the fee</td>
</tr>
<tr>
<td></td>
<td>Demo version</td>
</tr>
<tr>
<td></td>
<td>Presenting packages with different possibilities and fees</td>
</tr>
</tbody>
</table>
Result

“Automatic archive deletes within a specified amount of time, ”restoring deleted data,” open archives for public access,” getting certificates from public and private institutions,” “getting the complete version after paying the fee,” and “the demo version” were the features that we could eliminate. However, checking the current data extraction software shows that some of these features are of great importance.

Experts added “local storage” to the final checklist of the features.
ICA
ROMA 2022

September 19th-23rd, Italy

ARCHIVES: BRIDGING THE GAP
ARCHIVES: COMBLER LES FOSSES

9th Annual Conference of the International Council on Archives

19-23 September 2022 | Rome, Italy

Samaneh BORJI
borji.samaneh@gmail.com