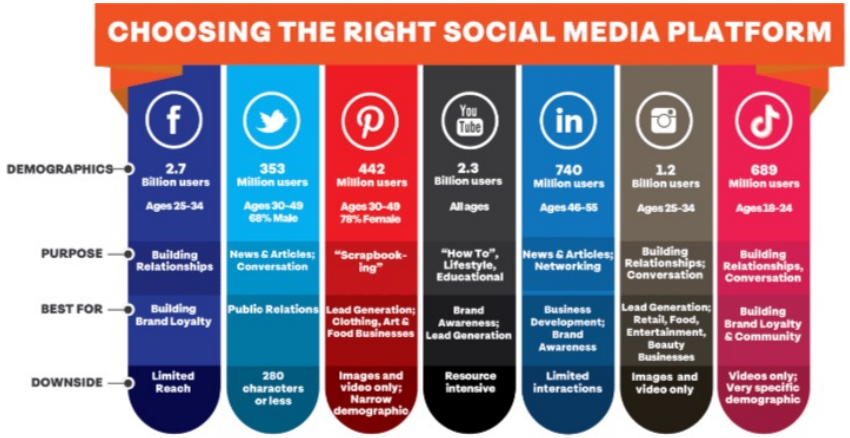
CoHUB
Cross-platform Bookmark Management App

Chen Sun

CONTENT

1 Introduction

Information Overload



Diversification of information source

Inefficiency of existing information management systems

CoHub:

A content management mobile application help to manage content online through cross-platform CMS framework and transformer-based techniques.

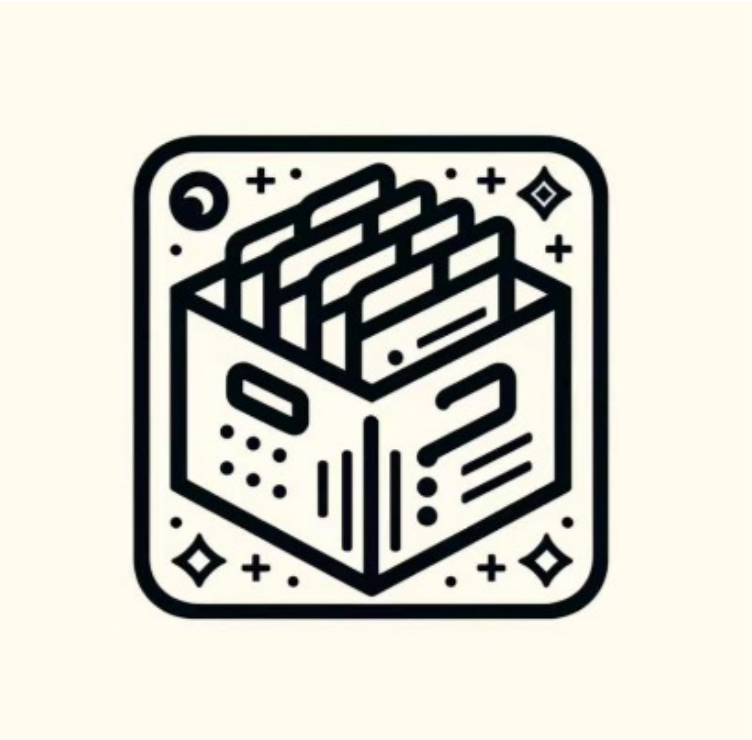


Figure: The logo of Cohub

Pocket

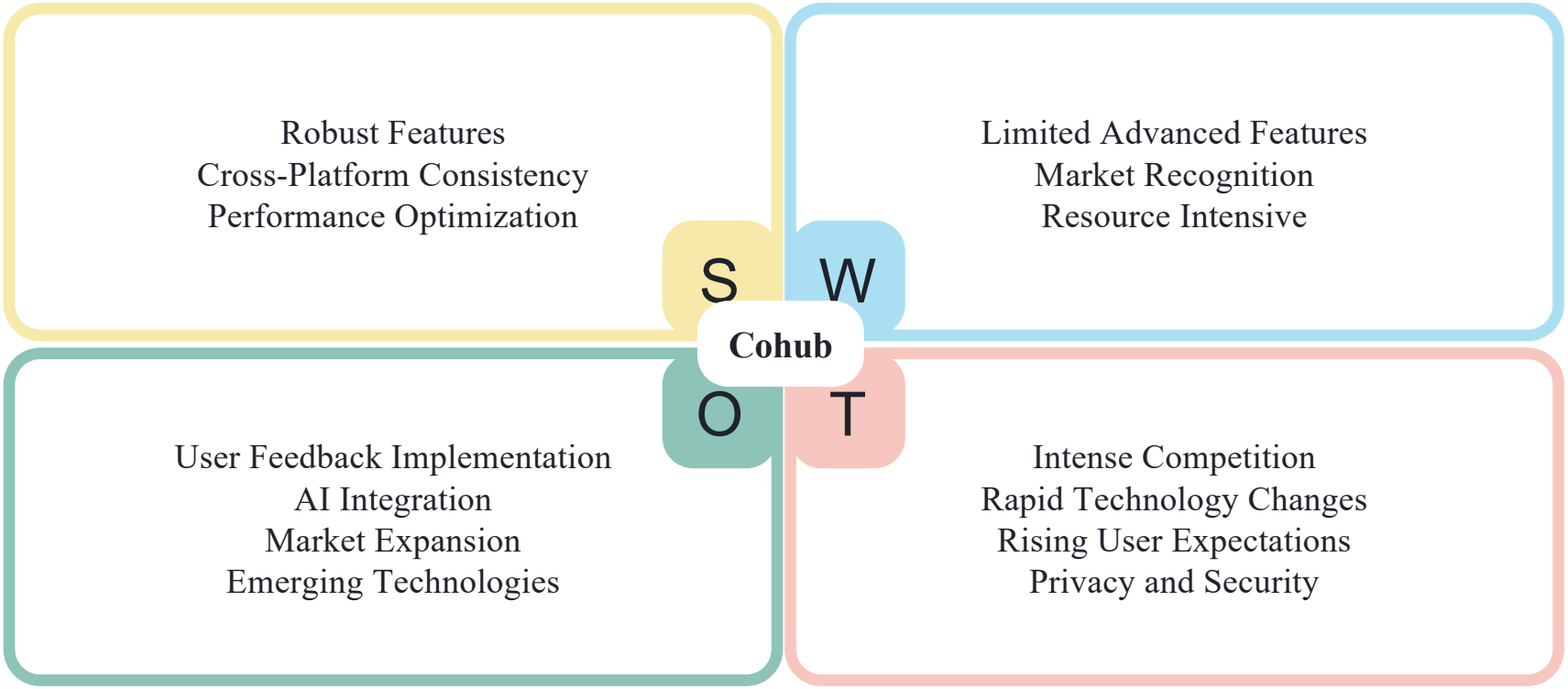
Pocket is an app for saving articles and videos to read or watch later, offering a distraction-free reading experience.

- Advantages: Simple interface for saving articles and videos.
- Disadvantages: Limited cross-platform support and link parsing functionality.

Evernote

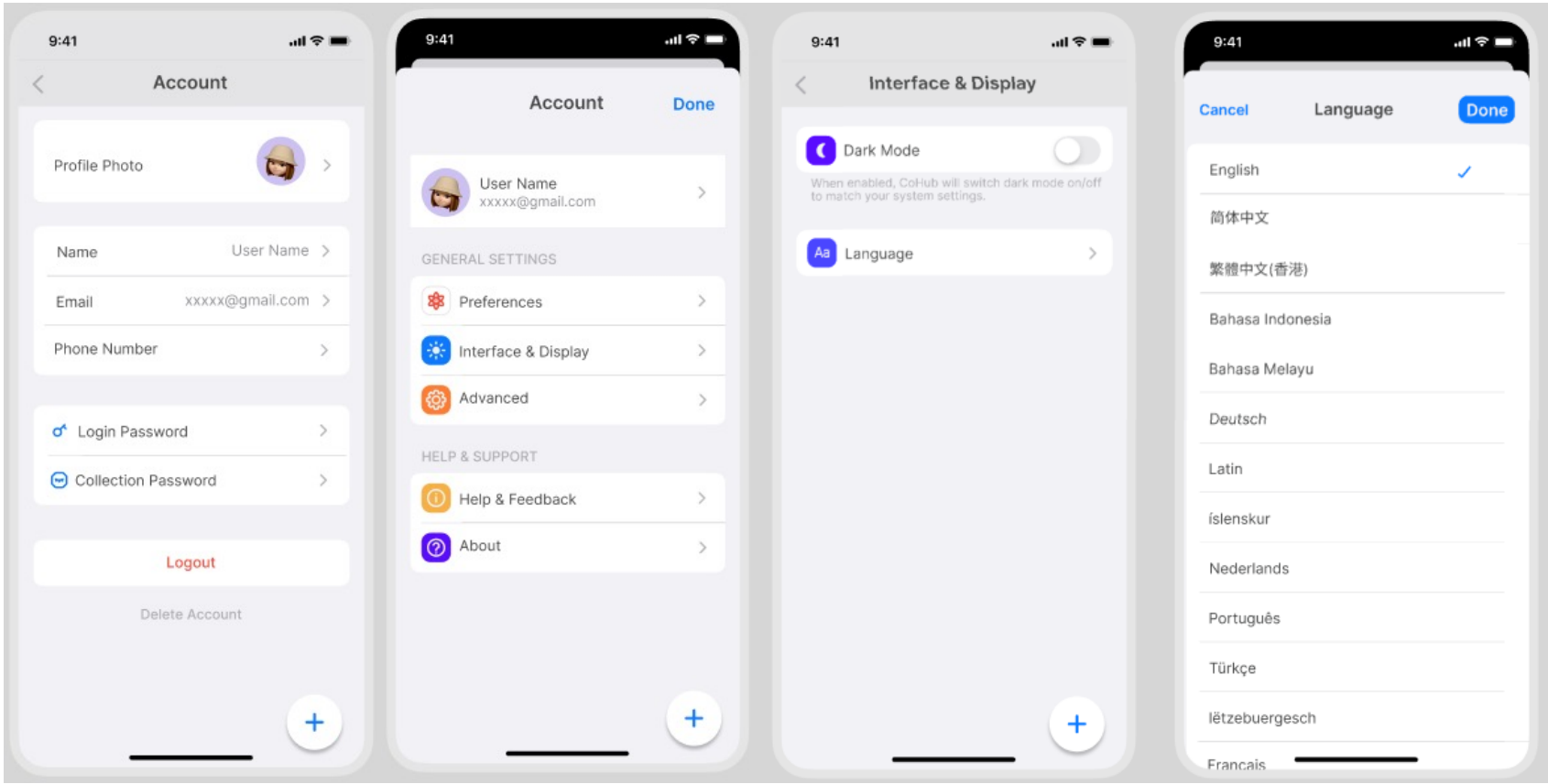
Evernote is a versatile note-taking and task management app that enhances productivity across devices.

- Advantages: Powerful document management and editing functions.
- Disadvantages: Weaker link parsing functionality and compatibility issues.


Robust Features
Cross-Platform Consistency
Performance Optimization
S
W
Limited Advanced Features
Market Recognition
Resource Intensive
Cohub
O
T
User Feedback Implementation
AI Integration
Market Expansion
Emerging Technologies
Intense Competition
Rapid Technology Changes
Rising User Expectations
Privacy and Security

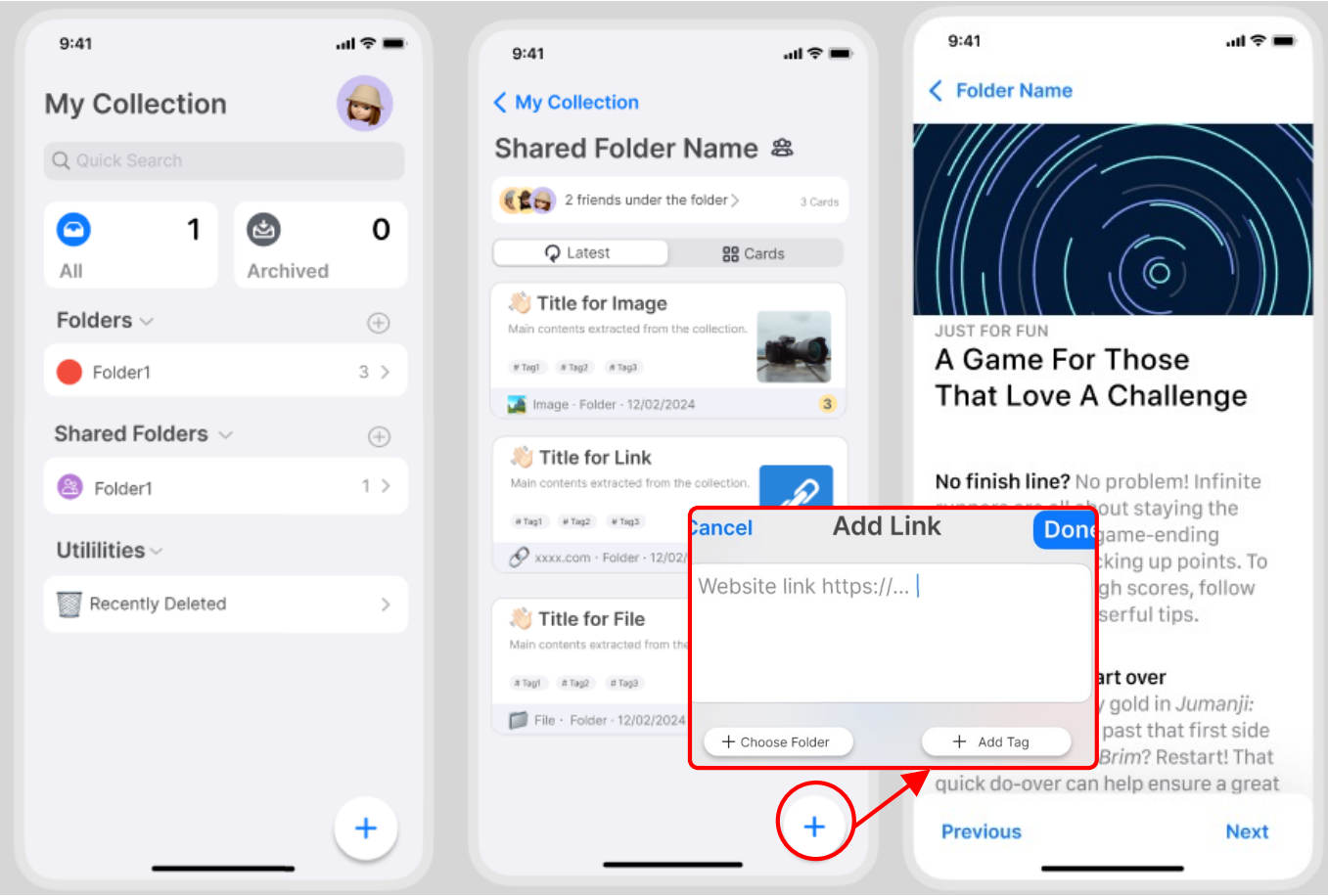
2 Functional Design

Account management and system settings



Content Collection

- **Personal Folder**
- **Shared Folder**
- **Archived Folder**
- **Recently Deleted Folder**

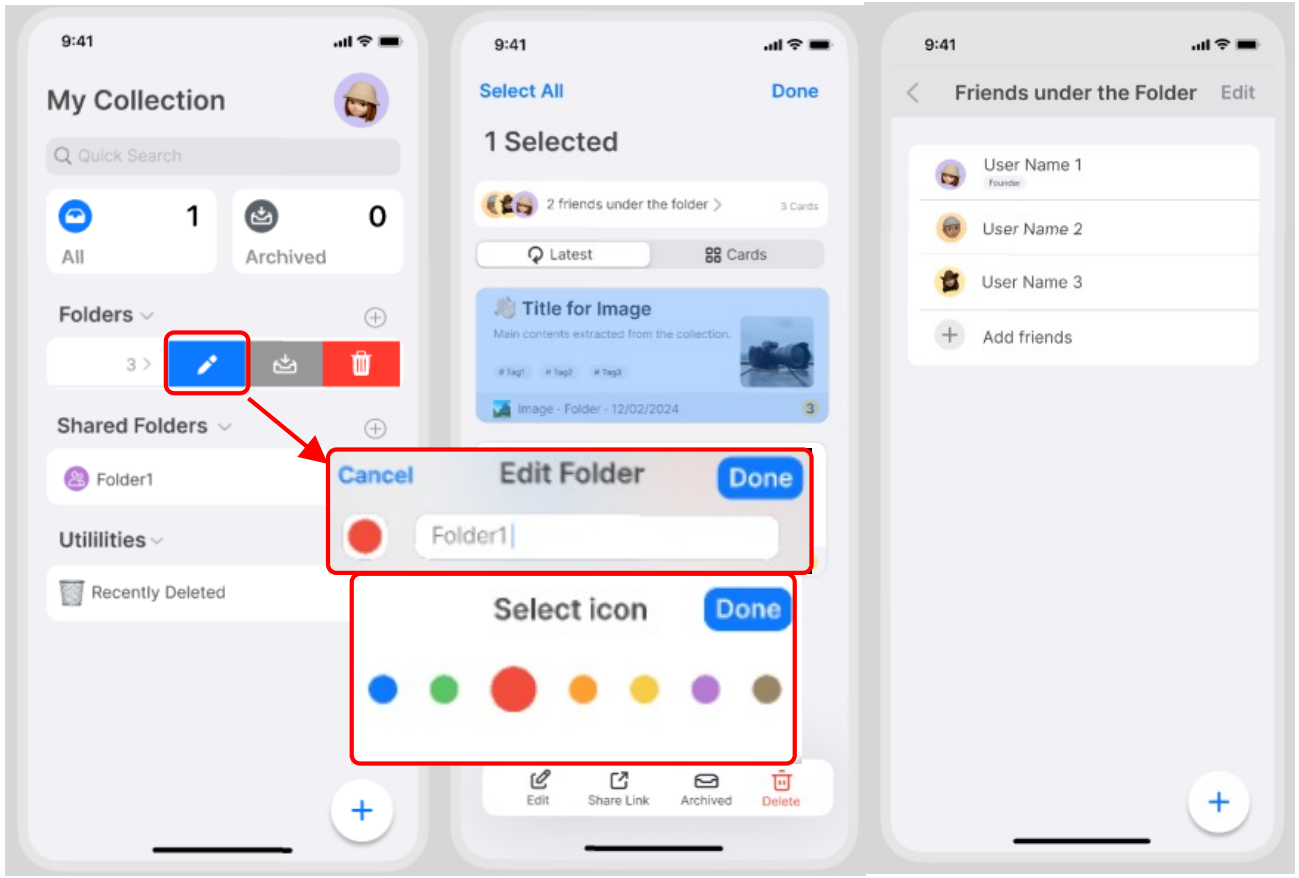


Content Management

1) **Personal Folder Management**

2) **Shared Folder Management**

- **Edit**
- **Share link**
- **Archived**
- **Delete**
- **...**



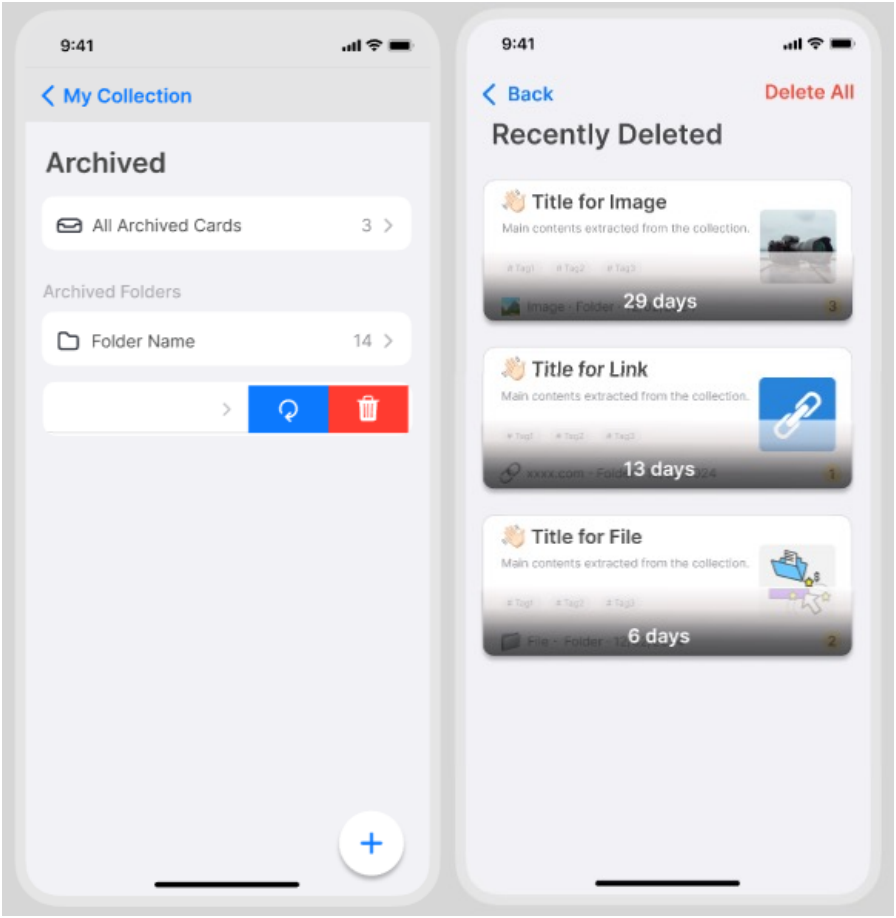
Content Management

3) Archived Folder Management

4) Recently Deleted Folder Management

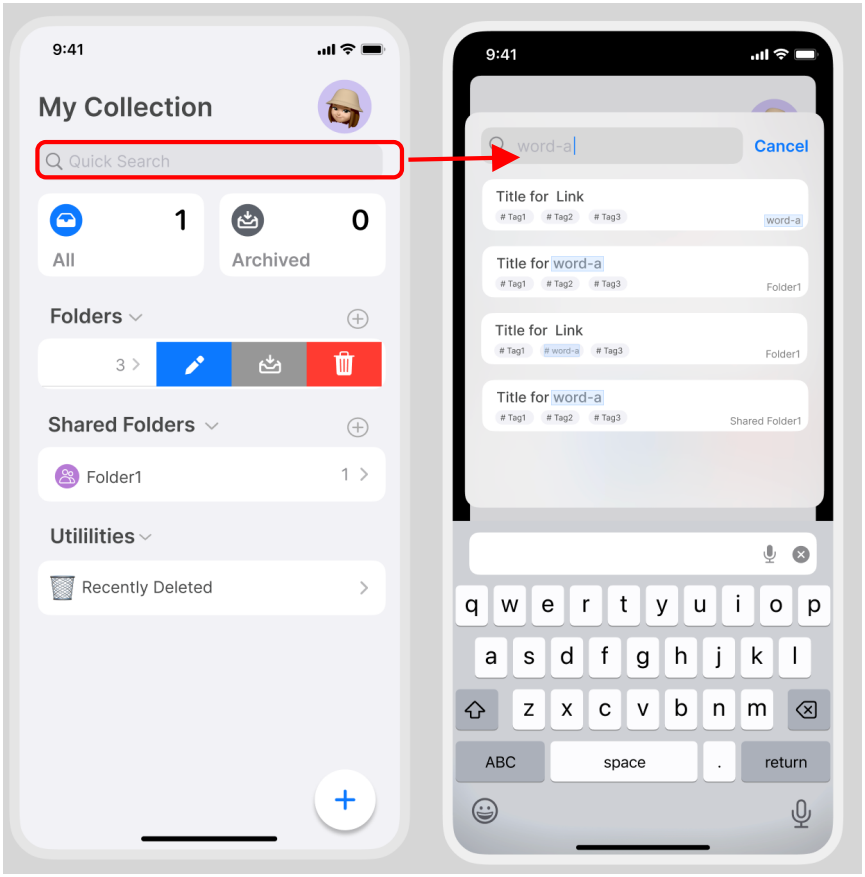
- **Recover**
- **Delete**
- **...**

The difference between archiving and deleting:
-Archived cards can be edited, viewed, searched, or recovered.
-Deleted cards are temporarily retained for 30 days and can be restored to folders.



Search Model

- Tag Search
  - Tag Search of Folder
  - Tag Search of Collected Cards
- Folder Search
  - Personal Folder
  - Shared Folder
  - Archived Folder
  - Recently Deleted Folder



3 Application Development

Overview of front-end development technologies

- **Architecture:** Front-end and back-end separation
- **Focus:** Mobile devices (Android, IOS)
- **Technologies:** Uni-APP, HBuilderX, Vue.js

Table 1: Hardware Environment

Hardware	Configuration
Operating system	Windows 11
CPU	13th Gen Intel(R) Core(TM) i5-13500H@2.60 GHz
Memory	32GB

Table 2: Software Environment

Software	Configuration
Programming Language	JavaScript, HTML, CSS
Framework	Vue.js, Uni-APP
Code Compilation Tool	HbuilderX
Debugging Tool	Customized IOS APP Base

The Structure of Page Layout

Main navigation bars: main pages and settings pages

- Main page: search, folder area, account settings, collection content, etc.
- Settings page: account and app settings.

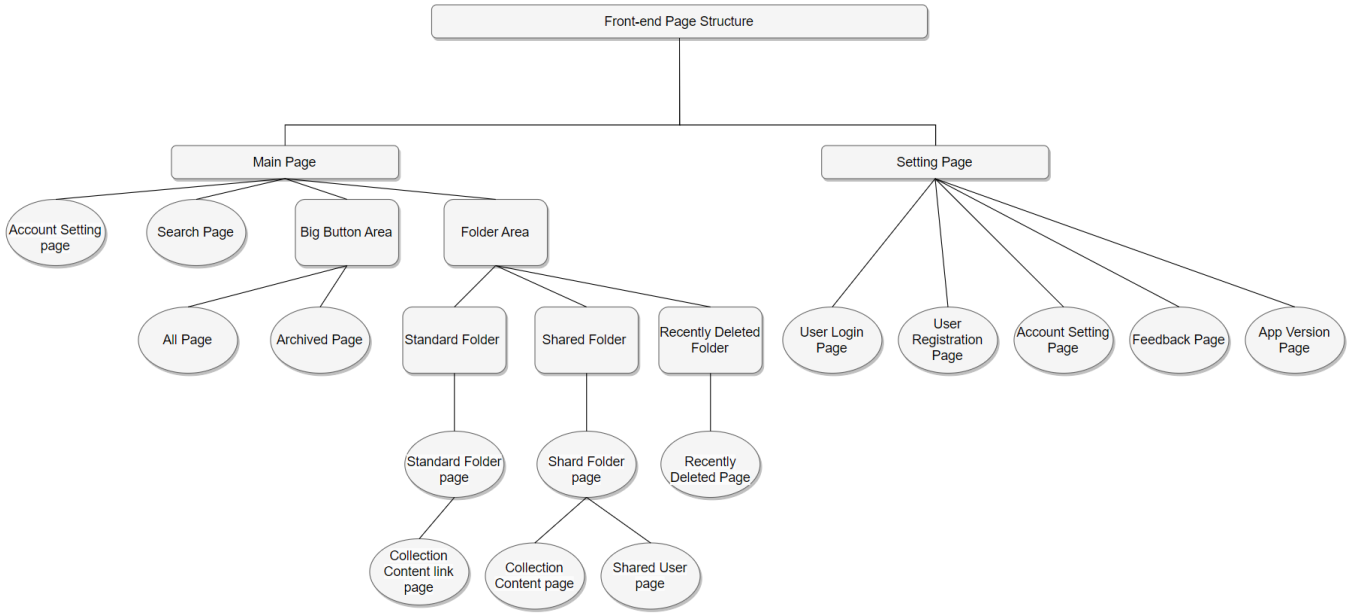


Figure: The Structure of Page Layout

Main Page Overview

- Search bar
- Navigation buttons
  - All / Archived information button
  - Count of collected/archived items
- Folder area
  - Personal / Shared / Deleted Folder
  - Folder item interactions: clicking / swiping

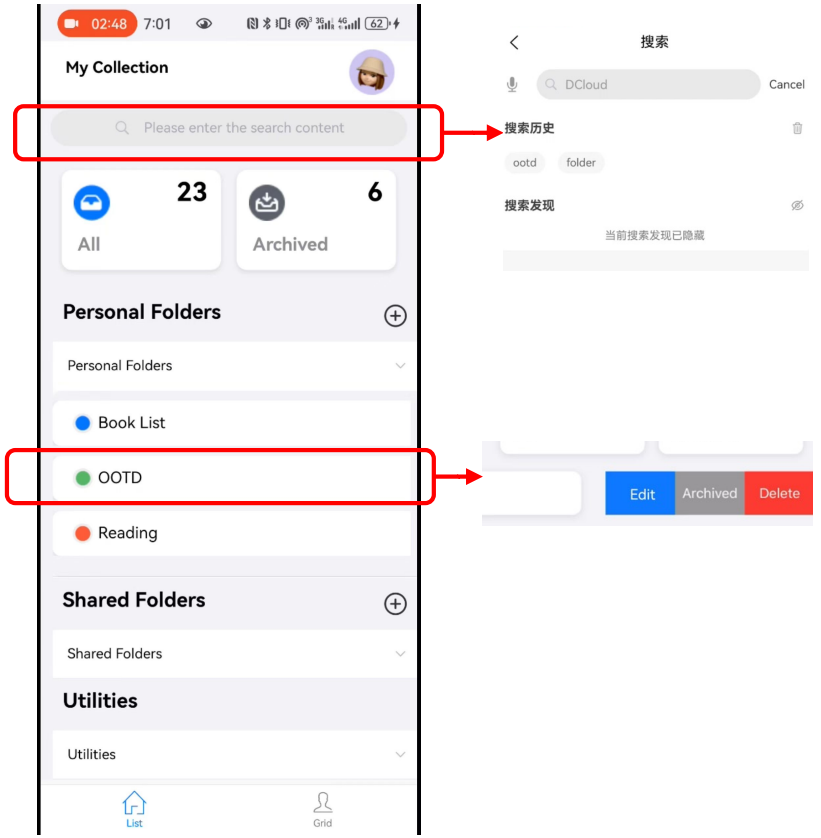


Figure: Main Page

Content Collection Page

Personal Folder:

- Displayed Type: Cards
- Elements: content, description, label, date, etc.
- Sorting options:
  - chronological/reverse order
  - simple mode
- Add the collection content

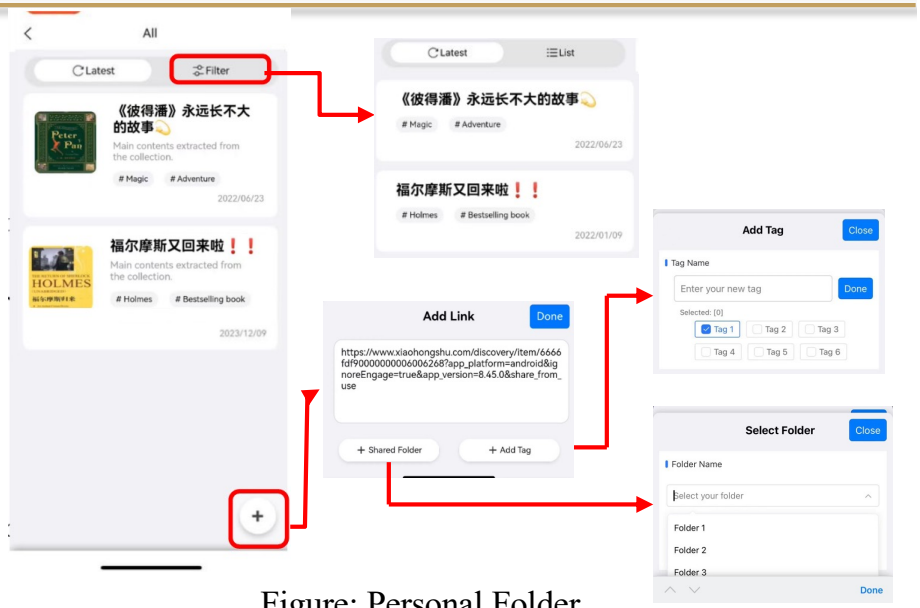


Figure: Personal Folder

Shared Folder:

- Additional shared user bar

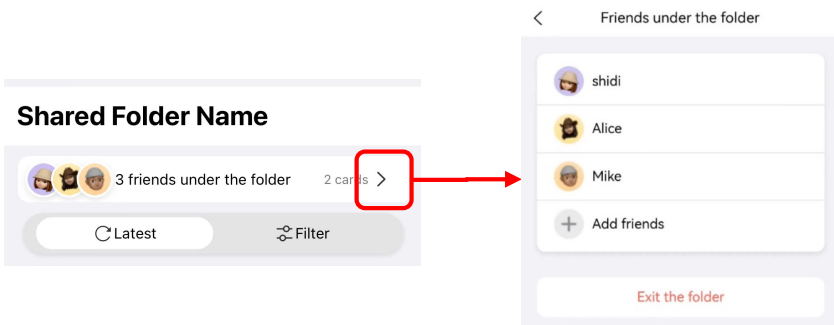


Figure: Shared Folder

Setting Page

- User management: personal information and account settings
- Features developed:
  - Account login
  - Language settings
  - App Introduction page
  - …

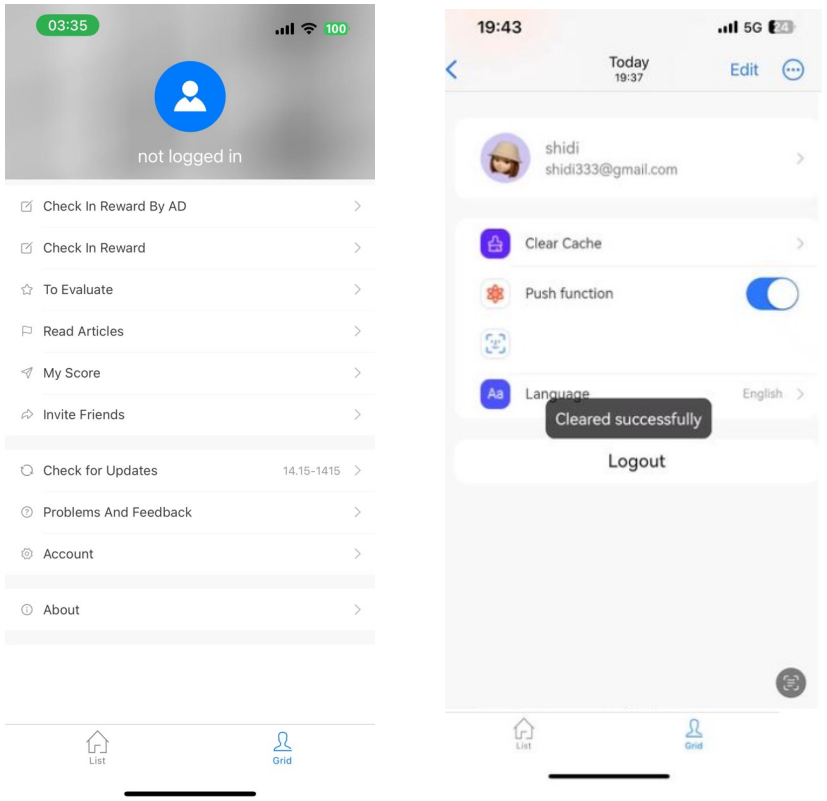


Figure: Setting Page

Account Registration

- Account registration:
  - Name
  - Password
  - Verification code
- Information change after registration:
  - Avatar
  - Nickname
  - Mobile number binding

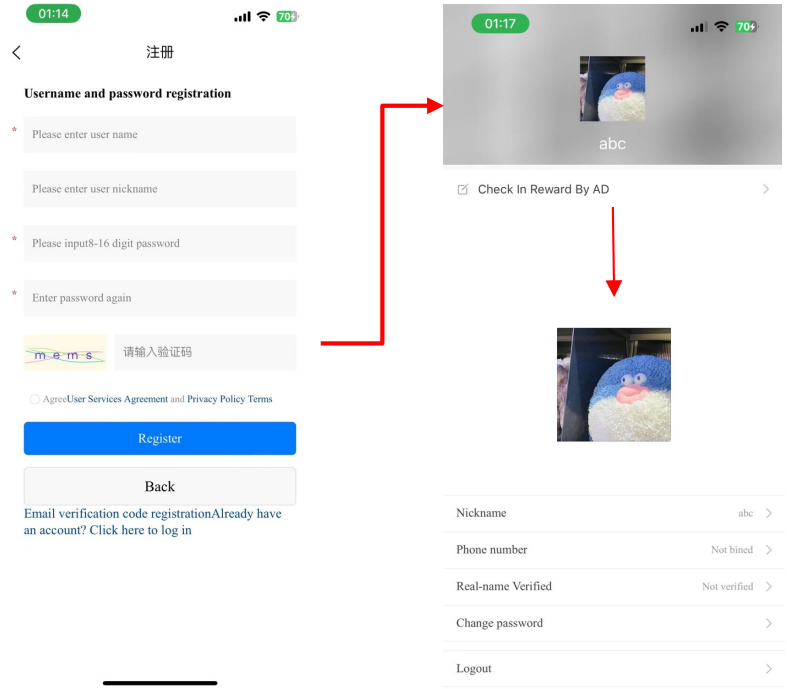


Figure: Account Registration Page

Account Login

- **Account Login Methods:**
  - Account Password Login
  - Password-free Login (SMS, WeChat, Apple login, etc.)
- Authentication: Uses third-party application Token authentication

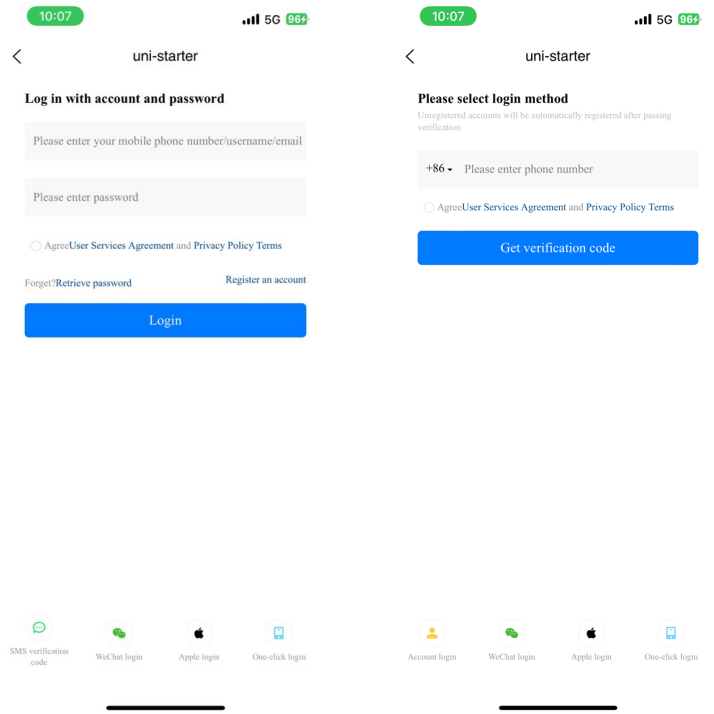
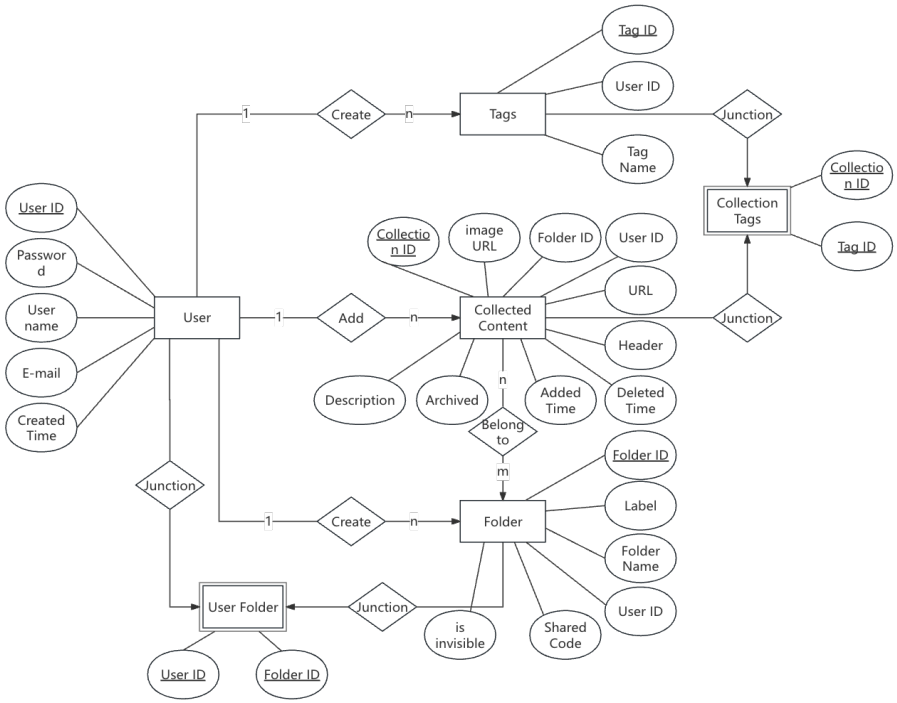


Figure: Account Password Login | Password-free Login

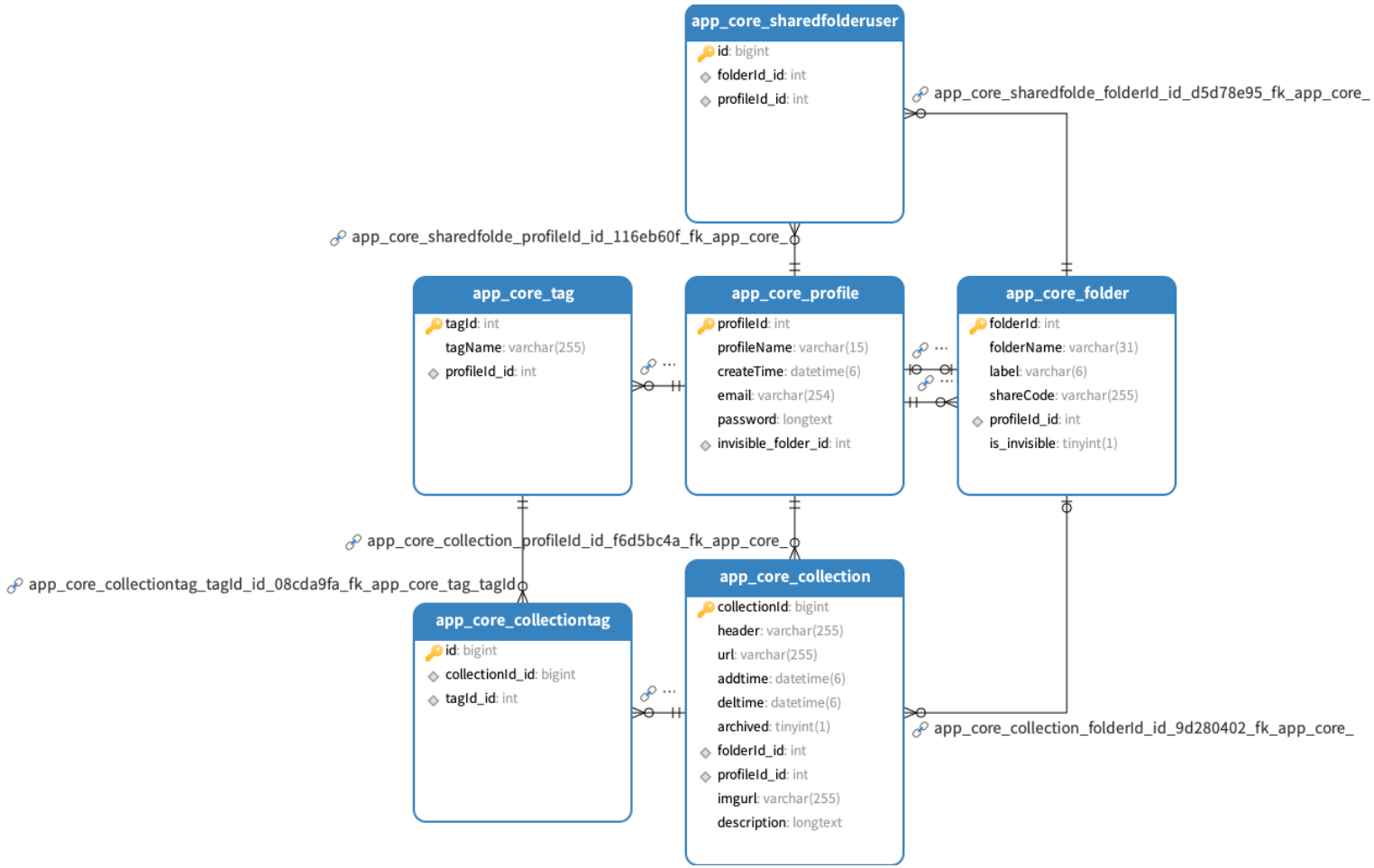


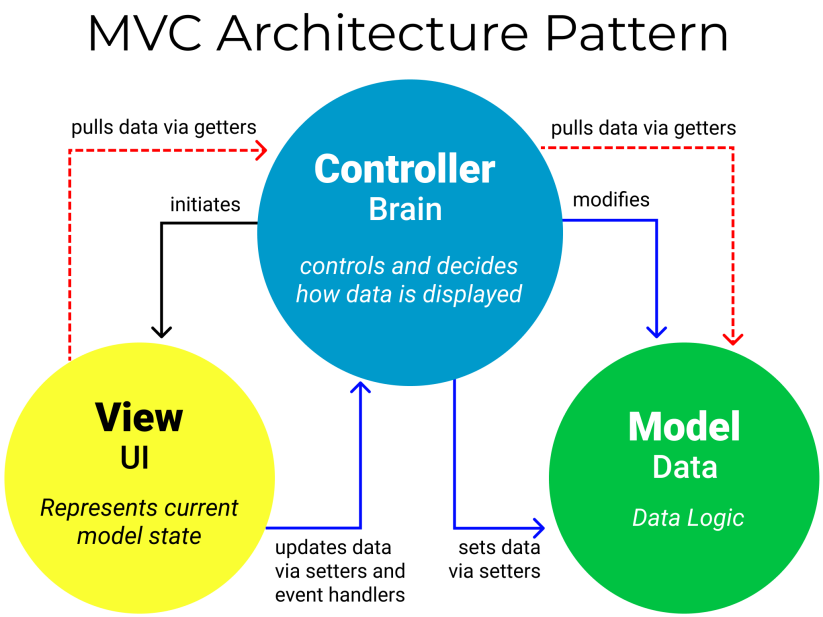
Database Design and Construction

- **Single Table Design**: In terms of database form design, a single table design was chosen. Put all users' data of the same type in the same form. This design approach has the characteristics of strong flexibility, easy maintenance, and high query efficiency

- **Association Table**: Using association tables to reflect the many to many relationship between two entities can make database operations easier and more conducive to database maintenance.

- **E-R Diagram**: The E-R diagram is shown on the left.


app_core_sharedfolderuser
id: bigint
folderId_id: int
profileId_id: int
app_core_sharedfolde_folderId_id_d5d78e95_fk_app_core_
app_core_sharedfolde_profileId_id_116eb60f_fk_app_core_
app_core_tag
tagId: int
tagName: varchar(255)
profileId_id: int
app_core_profile
profileId: int
profileName: varchar(15)
createTime: datetime(6)
email: varchar(254)
password: longtext
invisible_folder_id: int
app_core_folder
folderId: int
folderName: varchar(31)
label: varchar(6)
shareCode: varchar(255)
profileId_id: int
is_invisible: tinyint(1)
app_core_collection_profileId_id_f6d5bc4a_fk_app_core_
app_core_collectiontag_tagId_id_08cda9fa_fk_app_core_tag_tagId
app_core_collectiontag
id: bigint
collectionId_id: bigint
tagId_id: int
app_core_collection
collectionId: bigint
header: varchar(255)
url: varchar(255)
addtime: datetime(6)
deltime: datetime(6)
archived: tinyint(1)
folderId_id: int
profileId_id: int
imgurl: varchar(255)
description: longtext
app_core_collection_folderId_id_9d280402_fk_app_core_

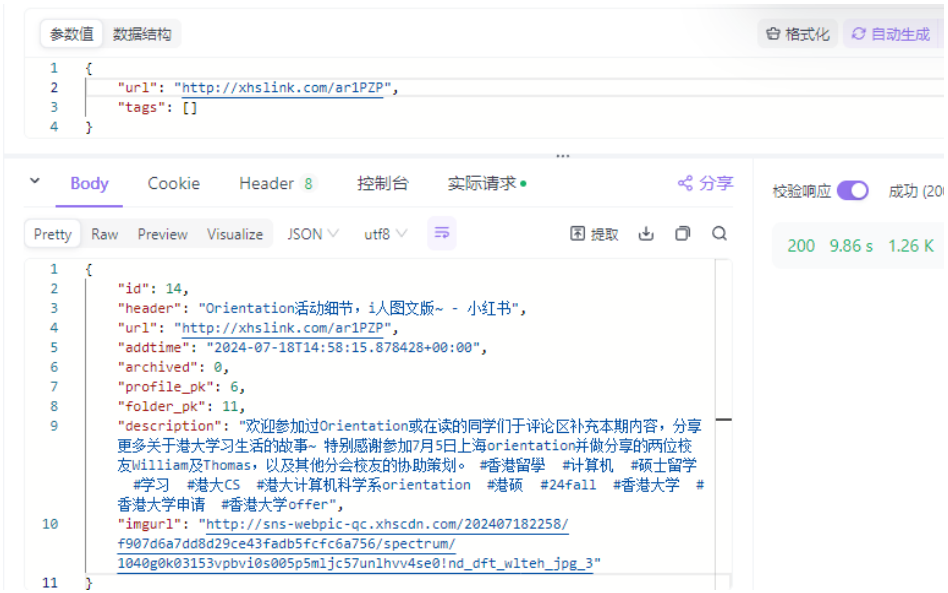


Back-end Technology

- **Django Frame**: Django is an advanced Python web framework that follows the MVC design pattern to support the rapid development and maintenance of complex data-driven websites.CoHub uses its built-in ORM, middleware support, and other features to implement back-end services.

- **RESTful API**: The HTTP protocol-based web service design approach designs interfaces that follow the Representational State Transfer (REST) principles. The communication between the front and back end is realized by transferring JSON.

- **Interface Test**: Test cases were designed using Apifox and interface test was completed to ensure that each function Api responded properly.



Key Technology : URL Link Parsing

- **BeatifulSoup**: BeautifulSoup is a Python library that can be used to parse HTML and XML documents. It creates a parse tree that makes it easy to extract information such as tags, text, attributes, and more. Here are the basic steps on how to extract information from a web page using the BeautifulSoup library:

- **Valuable Information Extraction**: Cohub extracts valuable, renderable information from web pages through the BeautifulSoup library. For example, the page title, main content and image links. This can be displayed in the front-end.

4 Algorithm Implementation

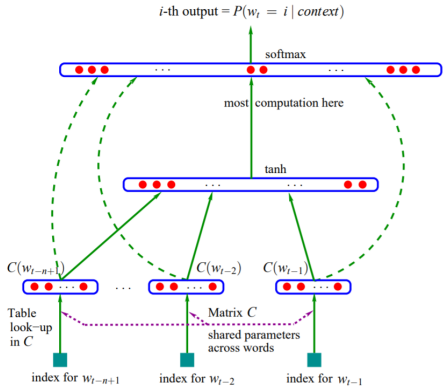
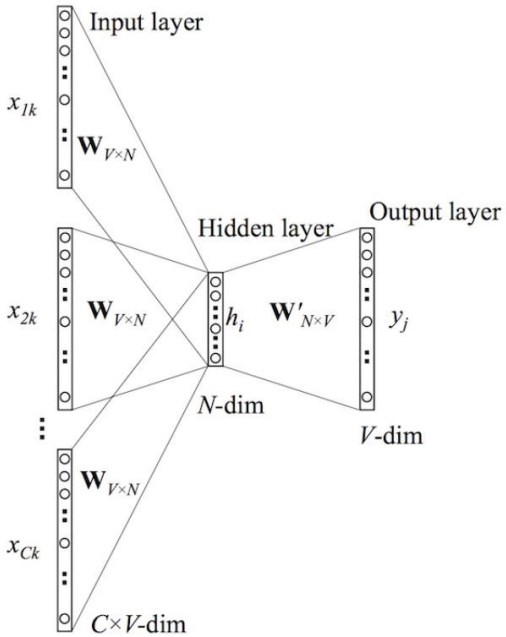
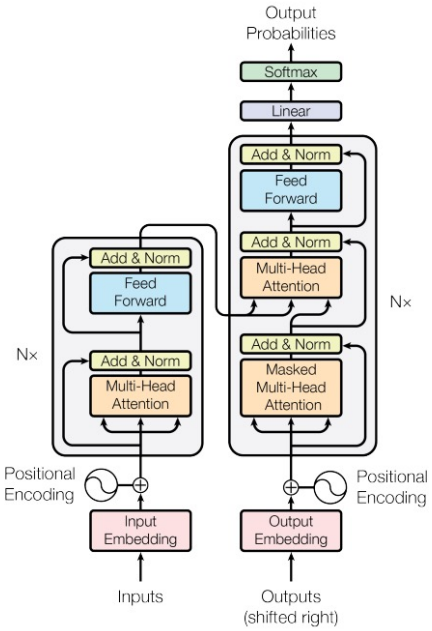
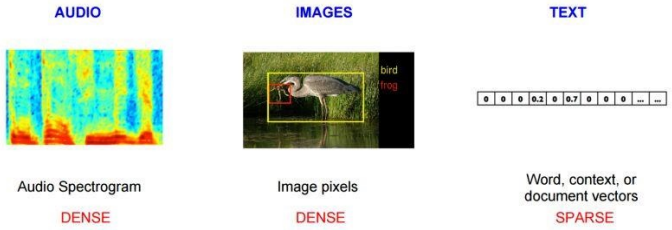
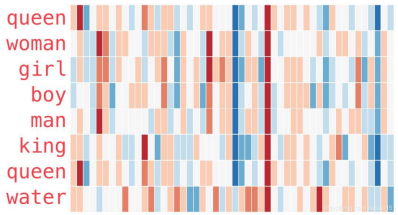


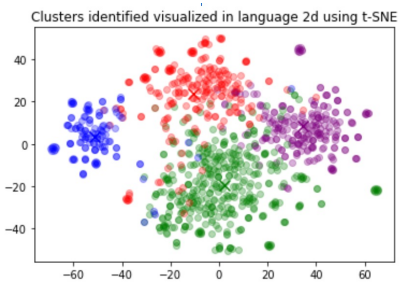
Figure 1: Neural architecture: $f(i, w_{t-1}, \cdots, w_{t-n+1}) = g(i, C(w_{t-1}), \cdots, C(w_{t-n+1}))$ where g is the neural network and $C(i)$ is the i-th word feature vector.



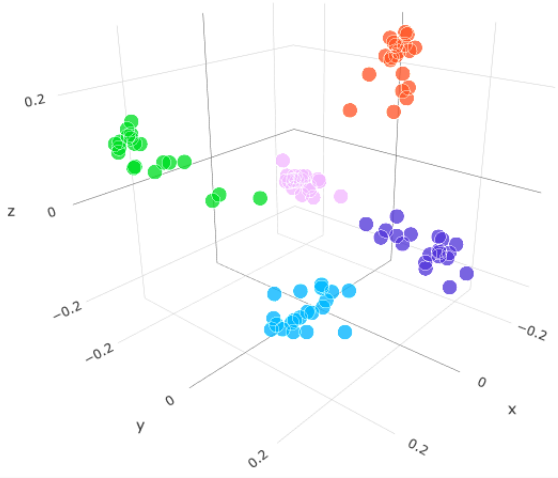






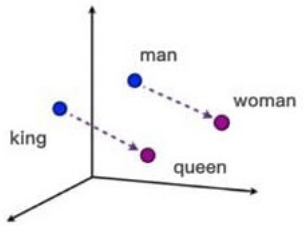


Problem: Tags are defined by users
Solution: Unsupervised Learning

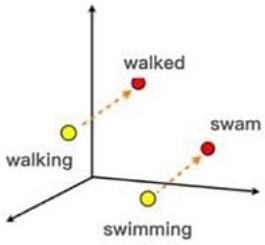


Reasons:

- **Incomplete tag for all collections**: Impratical for superviced learning
- **Effectiveness of Embeddings**: Semantic level classification
- **Applicability of K-means Clustering**: K-means is a widely used clustering algorithm that partitions data into K clusters while data points in different clusters are dissimilar.

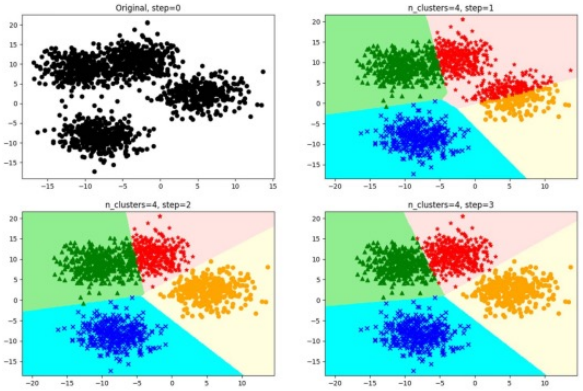


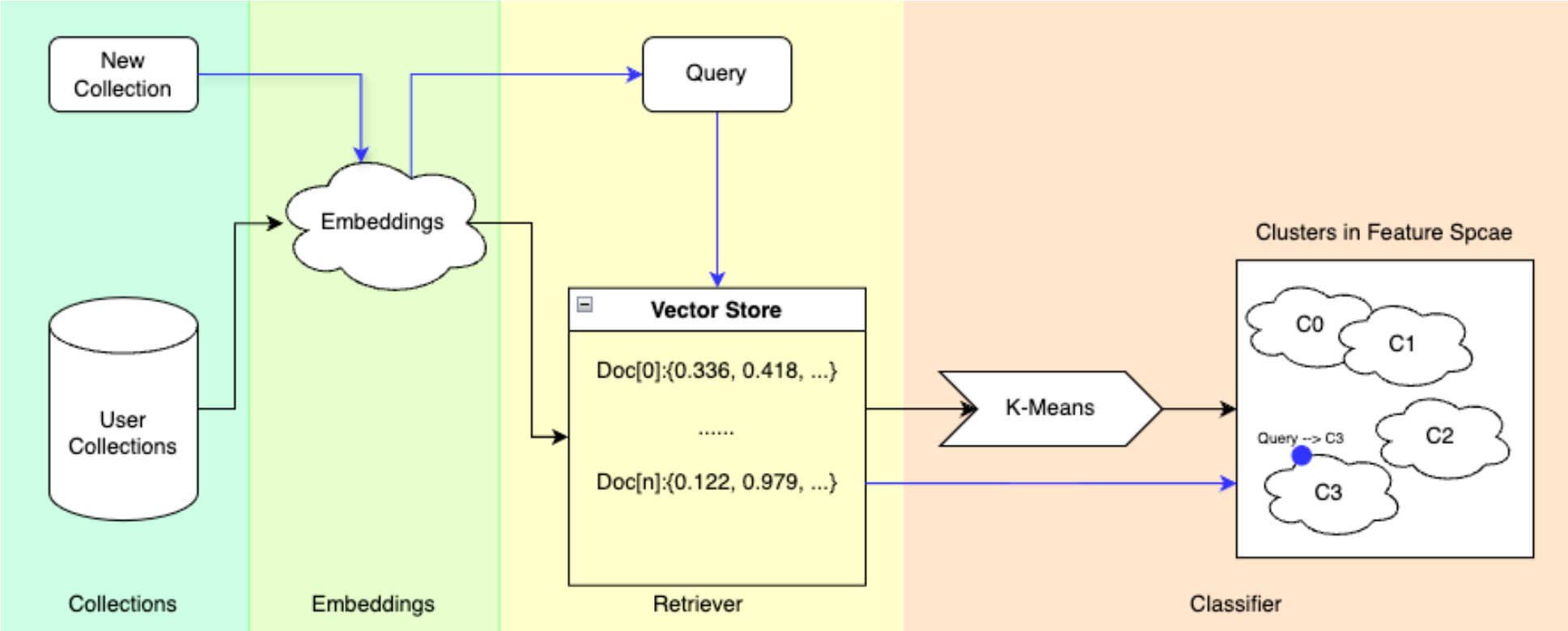
Male-Female

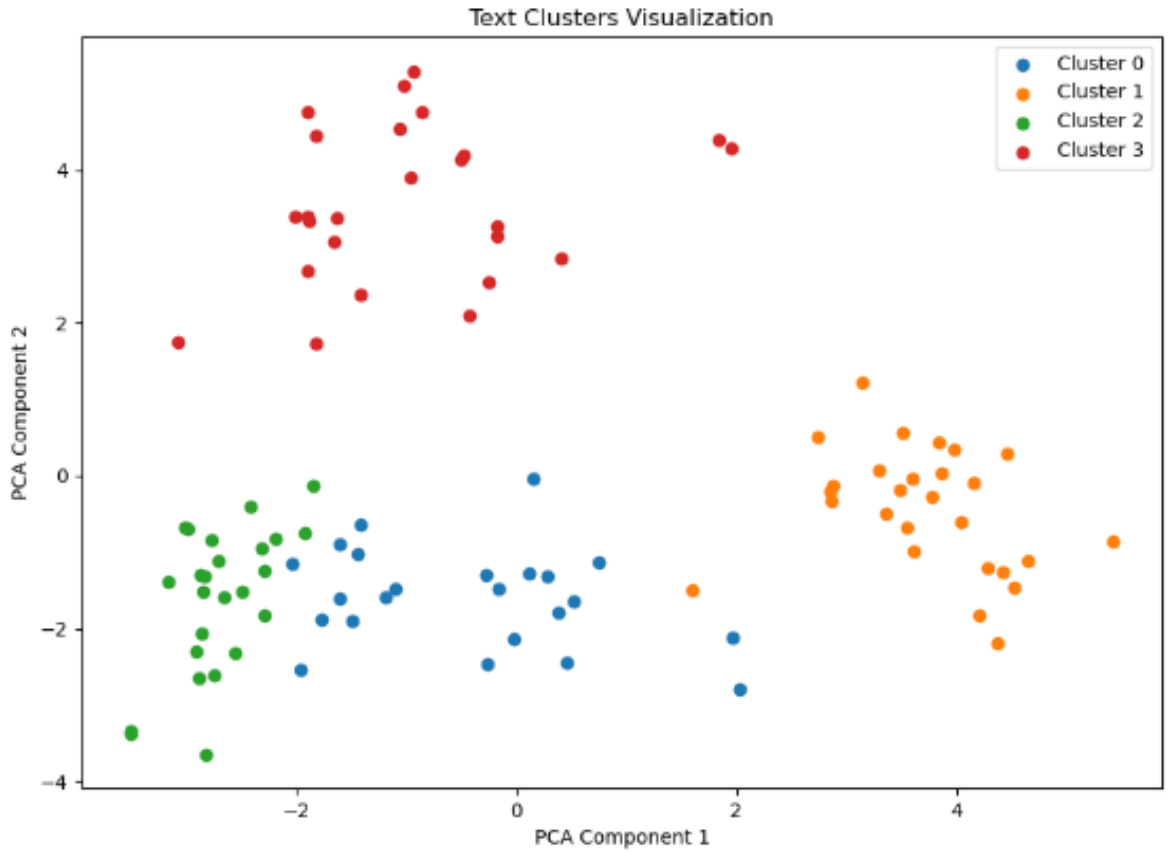


Verb tense

+




New Collection
Query
Embeddings
User Collections
Vector Store
Doc[0]:{0.336, 0.418, ...}
.....
Doc[n]:{0.122, 0.979, ...}
K-Means
Clusters in Feature Spcae
C0
C1
C2
C3
Query --> C3
Collections
Embeddings
Retriever
Classifier



- **Embeddings**: bge-zh-base-v1.5
- **Retrieve algorithm**:
  - Cosine Similarity (best)
  - KV-search (normal)
  - MMR (unacceptable)

- **Topic**: Sport, Trip, Cuisine, Pet
- **Accuracy**: 95/100
- **Error only occured on decision boundary**
  - Inherent advantage of K-Means
  - Good for decentralized centroids