

Zoho Email Contact Extractor Documentation

Overview

This Python script extracts email contacts from your Zoho Mail inbox, including sender names, subjects, timestamps, and attachments. It saves the data in Excel, JSON, and CSV formats with comprehensive analytics.

Key Features

- OAuth2 authentication
- Email contact extraction with metadata
- Attachment downloading (PDF, DOCX, XLSX, etc.)
- Rate limiting to avoid API bans
- Progress tracking and error handling
- Excel reports with summary statistics
- Domain-level analytics

Technologies Used

1. Core Libraries:

- *requests* - API interactions
- *pandas* - Data processing and Excel export
- *openpyxl* - Excel file operations
- *logging* - Error tracking and monitoring

2. APIs:

Zoho Mail API (messages, folders, accounts endpoints)
Zoho OAuth2 Authentication API

3. Infrastructure:

Local HTTP server for OAuth callback
Environment variables for secure credential management

Setup Instructions

1. Prerequisites:

- Python 3.7+
- Zoho Mail account
- [Zoho Developer Console](https://api-console.zoho.in/) access

2. Environment Setup:

Bash -

Clone repository

```
git clone https://github.com/your-repo/zoho-email-extractor.git
```

```
cd zoho-email-extractor
```

Install dependencies

```
pip install -r requirements.txt
```

3. Zoho API Configuration:

1. Register new OAuth client in [Zoho Developer Console](https://api-console.zoho.in/)
2. Add `http://localhost:5000/oauth/callback` as redirect URI
3. Note your `Client ID` and `Client Secret`

4. Configure Environment Variables:

Bash -

Linux/macOS

```
export ZOHO_CLIENT_ID="your_client_id"
```

```
export ZOHO_CLIENT_SECRET="your_client_secret"
```

Windows (PowerShell)

```
$env:ZOHO_CLIENT_ID="your_client_id"
```

```
$env:ZOHO_CLIENT_SECRET="your_client_secret"
```

Usage

Bash -

```
python zoho_email_extractor.py
```

Authentication Flow:

1. Script opens browser for Zoho login
2. Grant permission to access your mailbox
3. Authorization completes automatically
4. Extraction begins in terminal

Command-Line Options:

- `--max-messages` - Limit processed emails (default: 5000)
- `--no-attachments` - Skip attachment downloads
- `--batch-size` - Messages per API call (default: 50)

Output Files

All files save to `zoho_email_extraction/` directory:

1. Excel Report (`contacts_latest.xlsx`):

- Email Contacts Sheet
- Summary Statistics
- Domain Analytics
- Sender Frequency Rankings

2. JSON Export (`contacts_latest.json`):

- Full metadata including timestamps
- Attachment details with local paths

3. CSV File (`contacts_latest.csv`):

- Simplified contact list
- Compatible with email marketing tools

4. Attachments:

- Organized in `attachments/sender_email/` folders
- Filtered by file type (PDF, DOCX, etc.)

Key Implementation Details

1. OAuth Workflow:

- Local HTTP server handles redirect
- Access/refresh tokens stored securely
- Automatic token refresh before expiration

2. Rate Limiting:

- 40 requests/minute (Zoho API limit)
- Exponential backoff on 429 errors
- 1.2s delay between requests

3. Data Processing:

- Email deduplication with message counting
- Name extraction from email headers
- Timestamp normalization
- Domain segmentation

4. Attachment Handling:

- 10MB size limit per file
- Extension allowlist (PDF, DOCX, etc.)
- Sender-specific folders

Troubleshooting

Common Issues:

1. *Missing environment variables:*
 - Verify `.env` file or system variables
 - Re-export credentials
2. *401 Unauthorized errors:*
 - Delete ``tokens.json`` and re-authenticate
 - Check OAuth scope permissions
3. *Attachment download failures:*
 - Disable with ``--no-attachments`` flag
 - Verify Zoho API permissions include attachments
4. *Slow extraction:*
 - Reduce batch size (``--batch-size 25``)
 - Increase rate limit delay in code

Log Files:

Detailed logs in ``zoho_extractor.log``

Security Notes

- Never commit ``tokens.json`` to version control
- Attachments scanned for extensions only (no content validation)
- OAuth tokens expire automatically after 1 hour
- All sensitive data stored in memory only

Customization

Modify these constants in the script for different use cases:

Python

Increase max processed emails

```
self.max_messages = 10000
```

Add new file extensions

```
self.allowed_extensions.add('.pptx')
```

Adjust rate limiting

```
self.requests_per_minute = 30
```

License

Open-source MIT License. Use at your own risk. Not affiliated with Zoho Corp.