

Twitter Saver 2000

Instrukcja obsługi

Instalacja

W opisanym procesie instalacji zakłada się, że będzie ona przeprowadzana w systemie Ubuntu z zainstalowanym systemem zarządzania bazą danych PostgreSQL. Instalacja programu składa się z następujących kroków:

1. Utworzenie bazy danych służącej do przechowywania tweetów oraz użytkownika, na rzecz którego będzie działać program. Nadanie użytkownikowi uprawnień do zapisu elementów.

```
CREATE DATABASE twitter  
createuser twitter_saver -p
```
2. Wypakowanie archiwum zip
3. Przygotowanie pliku konfiguracyjnego
4. (Opcjonalnie) Przygotowanie konfiguracji w `/lib/systemd/system` jeżeli program ma działać jako usługa (co umożliwia automatyczne restartowanie).

Plik konfiguracyjny

Programy przyjmują parametr `--config`, wskazujący ścieżkę do pliku konfiguracyjnego. Domyślnie stosowany jest plik `config.yml` znajdujący się w lokalizacji, z której został uruchomiony program. Plik jest zapisany w formacie YAML i przechowuje mapowanie z polami będącymi mapowaniami:

1. `db` - parametry połączenia z bazą:
 - a. `host` - adres bazy danych, np. 127.0.0.1
 - b. `port` - port, na którym działa baza danych, np. 5432
 - c. `user` - użytkownik w bazie, np. `twitter_saver`
 - d. `password` - hasło podanego użytkownika, np. `twittersaver2000`
 - e. `dbName` - nazwa bazy danych, w której będą przechowywane dane, np. `twitter`
 - f. `sslMode` - czy podczas łączenia z bazą danych powinno być używane połączenie ssl, np. `Disable`.
2. `web` - parametry interfejsu web:
 - a. `port` - port, na którym będzie działać serwer, np. 8080
 - b. `secret` - klucz prywatny służący do podpisywania identyfikatorów sesji. Powinien być losowy i okresowo zmieniany. Klucze krótsze niż 256 bitów są niezalecane.
 - c. `users` - ciąg kont użytkowników. Każde konto jest mapowaniem:
 - i. `username` - nazwa użytkownika
 - ii. `password` - hasło
3. `twitter` - klucze do API Twittera:
 - a. `consumerKey`
 - b. `consumerSecret`

- c. token
 - d. tokenSecret
- 4. json
 - a. all - zapisuje całą treść tweeta do bazy
 - b. fields - ciąg nazw pól tweeta, które zostaną zapisane w bazie
- 5. autoDeleteDays - liczba dni, po których upłygnięciu tweet powinien zostać automatycznie usunięty lub null w celu wyłączenia usuwania

Dla programu web wymagane są parametry db i web a dla twitter-saver db, twitter, json i autoDeleteDays. W przypadku, gdy zabraknie któregoś parametru, program wyświetli błąd i zakończy działanie.

Konfiguracja usługi

W celu uruchomienia programu jako usługa należy w katalogu /lib/systemd/system utworzyć plik <nazwa>.service a w jego treści zapisać:

```
[Unit]
```

```
Description=Test service
```

```
ConditionPathExists=<ścieżka do programu>
```

```
After=network.target
```

```
[Service]
```

```
Type=simple
```

```
User=twittersaver
```

```
Group=twittersaver
```

```
LimitNOFILE=1024
```

```
Restart=on-failure
```

```
RestartSec=10
```

```
WorkingDirectory=<ścieżka do katalogu zawierającego program>
```

```
ExecStart=<ścieżka do programu>
```

```
PermissionsStartOnly=true
```

```
ExecStartPre=/bin/mkdir -p /var/log/<nazwa>
```

```
ExecStartPre=/bin/chown syslog:adm /var/log/<nazwa>
```

```
ExecStartPre=/bin/chmod 755 /var/log/<nazwa>
```

```
StandardOutput=syslog
```

```
StandardError=syslog
```

```
SyslogIdentifier=test
```

```
[Install]
```

```
WantedBy=multi-user.target
```

Usługa będzie działała na rzecz użytkownika twittersaver (można podać innego). W parametrze ExecStart można dodać parametr --config definiujący ścieżkę do pliku konfiguracyjnego. Dla każdego programu należy utworzyć osobną usługę. Następnie należy

wykonać komendę `sudo systemctl enable <nazwa>.service`. W celu uruchomienia usługi można użyć polecenia `sudo systemctl start <nazwa>`, logi można podejrzeć przy pomocy polecenia `sudo journalctl -f -u <nazwa>`.

Uruchomienie

Należy uruchomić programy `web` oraz `twitter-saver`

Zarządzanie śledzonymi obiektami

Zarządzanie śledzonymi obiektami jest możliwe z wykorzystaniem interfejsu `web`, który jest dostępny na porcie zadanym w konfiguracji i chroniony za pomocą mechanizmu *JSON Web Tokens* (użytkownicy są pobierani z pliku konfiguracyjnego). Po wejściu na stronę, użytkownik proszony jest o podanie danych logowania. Po poprawnym logowaniu wyświetlony zostaje interfejs administratora `Twitter Saver`. Składa się on z 3 paneli (1 jest widoczny dopiero po wybraniu obiektu z listy):

- Nowy obiekt - umożliwia dodanie nowego obiektu do śledzenia.
- Śledzone obiekty - wyświetla listę śledzonych obiektów
- Zarządzaj obiektem - umożliwia wyświetlenie statystyk dotyczących obiektu oraz jego usunięcie. Wyświetlany dla obiektu wybranego z listy w panelu Śledzone obiekty (w szczególności nie jest wyświetlany, jeśli nic nie zostanie wybrane)

Nowy obiekt

W celu dodania nowego obiektu do śledzenia należy zdefiniować typ obiektu (użytkownik lub słowo kluczowe), podać w polu tekstowym poniżej odpowiednio nazwę śledzonego obiektu (czyli nazwę użytkownika lub słowo kluczowe) oraz wybrać tryb pobierania (stream dla pobierania strumieniowego i batch dla pobierania wsadowego).

W przypadku wybrania trybu wsadowego należy dodatkowo podać datę początkową pobierania historii. Pobrane zostaną tweety wysłane po północy zadanego dnia w strefie czasowej użytkownika.

Pobieranie strumienia zostanie rozpoczęte w ciągu minuty od dodania obiektu.

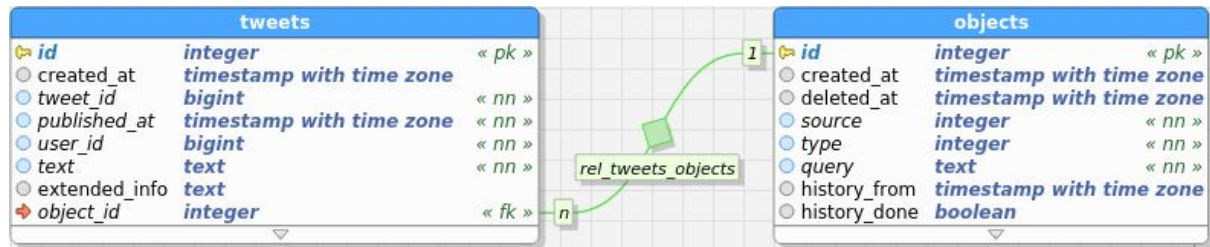
Zarządzaj obiektem

Po wybraniu obiektu z listy śledzonych obiektów w tym panelu wyświetlane są statystyki dotyczące całego śledzonego okresu dla tego obiektu. Liczba tweetów w poszczególnych dniach jest liczona zawsze w strefie czasowej serwera.

Na górze panelu występują pola `Od` i `Do` umożliwiające odfiltrowanie zakresu dat, dla którego mają zostać wyznaczone statystyki. Każdorazowo po zmianie którejkolwiek z tych dat wyświetlone zostaną statystyki dotyczące okresu zaczynającego się o północy dnia `Od` i kończącego się o północy dnia następującego po dniu `Do` w strefie czasowej użytkownika.

Po kliknięciu przycisku `Usuń obiekt` zostanie wyświetlony dialog potwierdzenia z możliwością zaznaczenia opcji `"Usuń permanentnie"`, która usunie również zebrane tweety. Jeżeli ta opcja nie zostanie zaznaczona, to obiekt jest jedynie oznaczany jako usunięty (nie będą pobierane nowe tweety). Usunięcie obiektu usuniętego usuwa go permanentnie.

Baza danych



Zebrane tweety są przechowywane w bazie danych zgodnej z powyższym schematem. Tabela **objects** służy do przechowywania obiektów (definiowanych poprzez interfejs web), a w tabeli **tweets** pojawiają się tweety ich dotyczące pobrane z API twittera.

Wspólne kolumny:

- **id** - automatycznie wygenerowany identyfikator rekordu,
- **created_at** - czas utworzenia rekordu,

Tabela **objects**:

- **deleted_at** - czas usunięcia rekordu,
- **source** - informacja czy obiekt pobiera dane z historii czy streama,
- **type** - rodzaj obiektu (użytkownik czy słowo kluczowe),
- **query** - nazwa śledzonego obiektu (nazwa użytkownika lub słowo),
- **history_from** - czas od którego ma zostać pobrana historia,
- **history_done** - informacja czy cała historia została pobrana

Tabela **tweets**:

- **tweet_id** - unikalny liczbowy identyfikator tweeta (nadany przez twitter),
- **published_at** - czas publikacji tweeta,
- **text** - pełny tekst tweeta,
- **user_id** - identyfikator użytkownika, który napisał tweeta,
- **extended_info** - zapisane w formacie JSON dodatkowe pola z odpowiedzi Twittera zdefiniowane w pliku konfiguracyjnym (mapowanie json).