

# RssHunter

---

Technologie i programowanie WWW



**Jakub Kozik i Dariusz Gruca**

**2012-06-20**

## Spis treści

Cel programu .....	3
Podstawowe funkcjonalności programu .....	3
Interfejs programu .....	4
Główne okno.....	4
Okna poboczne .....	5
Okno tworzenia nowej bazy .....	5
Okno parametrów bazy .....	5
Okno ustawień użytkownika.....	6
Przed rozpoczęciem.....	7
Szybkie wprowadzenie - przykładowy scenariusz .....	8
Ustawienie danych użytkownika .....	8
Dodanie nowej bazy przeszukiwań.....	9
Ustawienie parametrów bazy.....	9
Ustawienie sposobu i częstotliwości powiadomień .....	10
Zakończenie .....	11
Wykorzystane technologie i biblioteki .....	13
Diagram stanów .....	13
Baza danych.....	14
Struktura programu.....	14
Szczegóły działania programu .....	15
Tworzenie baz przeszukiwań .....	15
Odświeżanie bazy RSS.....	15
Odświeżanie bazy Html.....	15
Planowanie automatycznego odświeżania baz .....	15
Wysyłanie powiadomień mailowych .....	16
Wysyłanie powiadomień SMS .....	16

# Wstęp

---

## Cel programu

Program RssHunter służy do „polowania” na konkretne informacje w kanałach RSS. Użytkownik nie musi dzięki temu tracić czasu na przeszukiwanie wszystkich nowych informacji, jakie się pojawiają. Nie musi też martwić się o odświeżanie kanału w poszukiwaniu nowych wpisów. Jedyne, co robi to tworzy bazę poszukiwań, do której dodaje kanały RSS oraz słowa kluczowe, które powinny znajdować się w interesujących go artykułach. Jeśli tylko na którymś kanale pojawi się wpis, który użytkownika interesuje, program o tym powiadamia.

## Podstawowe funkcjonalności programu

Podstawowymi funkcjonalnościami programu są:

- ❖ Sprawdzanie czy w bazie RSS pojawiły się nowe, interesujące użytkownika wpisy. To, czy wpis jest poszukiwany przez użytkownika, jest stwierdzane na podstawie podanych przez niego słów kluczowych.
- ❖ Sprawdzanie czy zmieniła się zawartość jakiejś konkretnej strony www. Kiedy strona zostaje zaktualizowana (np. zostały do niej dodane nowe wpisy), użytkownik otrzymuje powiadomienie. W przypadku dynamicznych stron użytkownik może określić konkretne tagi html, których aktualizacja go interesuje.

Powiadomienia użytkownika realizowane są w następujący sposób:

- ❖ Email – użytkownik dostaje na podany adres emailowy powiadomienie o zmianie zawartości bazy poszukiwań. Do wiadomości dołączony zostaje krótki raport na temat aktualizacji bazy.
- ❖ Sms – użytkownik dostaje powiadomienie sms nowych informacjach w bazie. Funkcja jest przydatna, kiedy użytkownik nie ma dostępu do komputera, a czeka na jakąś ważną informację.
- ❖ Popup – kiedy użytkownik aktualnie korzysta z komputera, a nie chce zawracać sobie głowy sprawdzaniem czy na stronach z rssami pojawiło się coś, co go interesuje, może ustawić powiadomienia przez popup. W rogu ekranu pojawi się nieprzeszkadzająca w pracy chmurka,

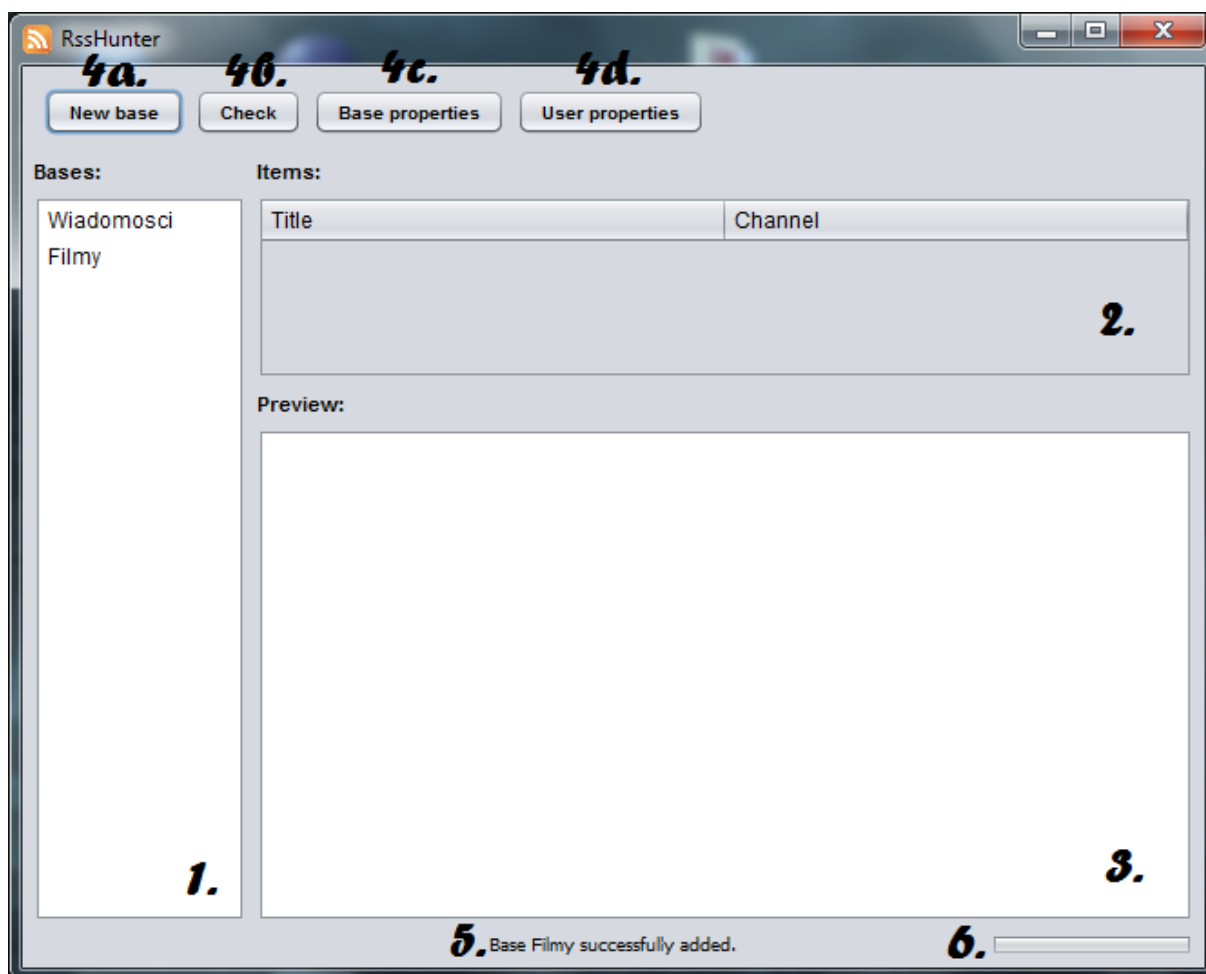
która powiadomi, że w bazie znajdują się nowe wpisy, które użytkownik może przejrzeć w wolnej chwili.

Program umożliwia także przeglądanie nowych wpisów w bazach. Poza panelem zarządzania bazami przeszukiwań dostępne są także panele z listą nowych wpisów oraz podglądem konkretnego wpisu. Bezpośrednio z podglądu, za pomocą jednego kliknięcia, użytkownik może otworzyć pełny wpis w przeglądarce internetowej.

## Interfejs programu

### Główne okno

Główne okno programu służy do zarządzania bazami przeszukiwań oraz do przeglądania nowych wpisów w tych bazach.



Rys 1. Okno główne programu.

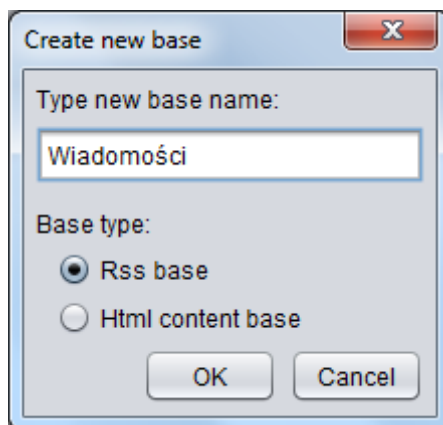
Interfejs składa się z następujących komponentów:

1. Lista z bazami przeszukiwań.
2. Lista z nowymi wpisami w bazie. Zawiera tytuł wpisu oraz kanał, z którego wpis pochodzi.
3. Podgląd wybranego wpisu.
4. Przyciski zarządzania bazami:
  - a) Tworzenie nowej bazy przeszukiwań
  - b) Wymuszenie sprawdzenia czy w bazie są nowe wpisy
  - c) Ustawienia parametrów wybranej bazy
  - d) Ustawienia użytkownika
5. Pasek stanu.
6. Pasek postępu sprawdzania bazy.

## Okna poboczne

### Okno tworzenia nowej bazy

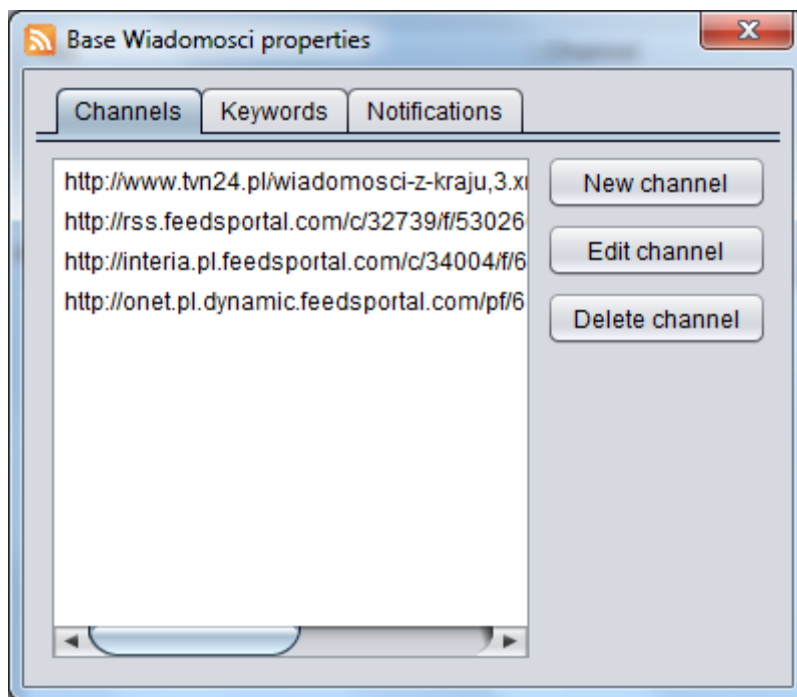
Należy podać nazwę bazy oraz jej rodzaj. Baza może pobierać wpisy z kanałów rss oraz sprawdzać czy na konkretnej stronie html zmieniła się zawartość.



Rys 2. Okno tworzenia nowej bazy.

### Okno parametrów bazy

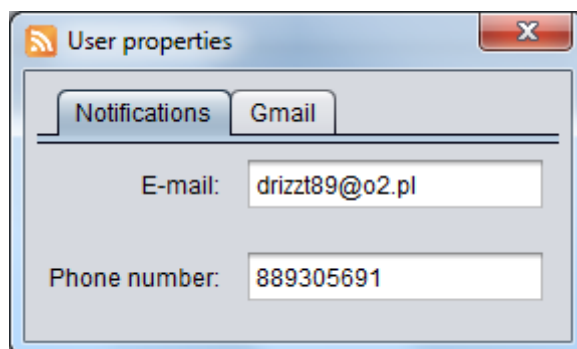
Pozwala na zarządzanie bazą przeszukiwań. Możemy do niej dodawać adresy kanałów/stron oraz słowa kluczowe/tagi. Ostatnia zakładka służy do ustawienia częstotliwości sprawdzania bazy oraz typu powiadomień.



Rys 3. Okno właściwości bazy.

### Okno ustawień użytkownika

Okno służy do ustawienia informacji o użytkowniku. W pierwszej zakładce ustawiany jest adres mailowy oraz telefoniczny, na który powinny przychodzić powiadomienia. W drugiej ustawiane są dane do konta gmail, które obsługuje wysyłanie smsów oraz maili.



Rys 4. Okno ustawień użytkownika.

# Poradnik użytkownika

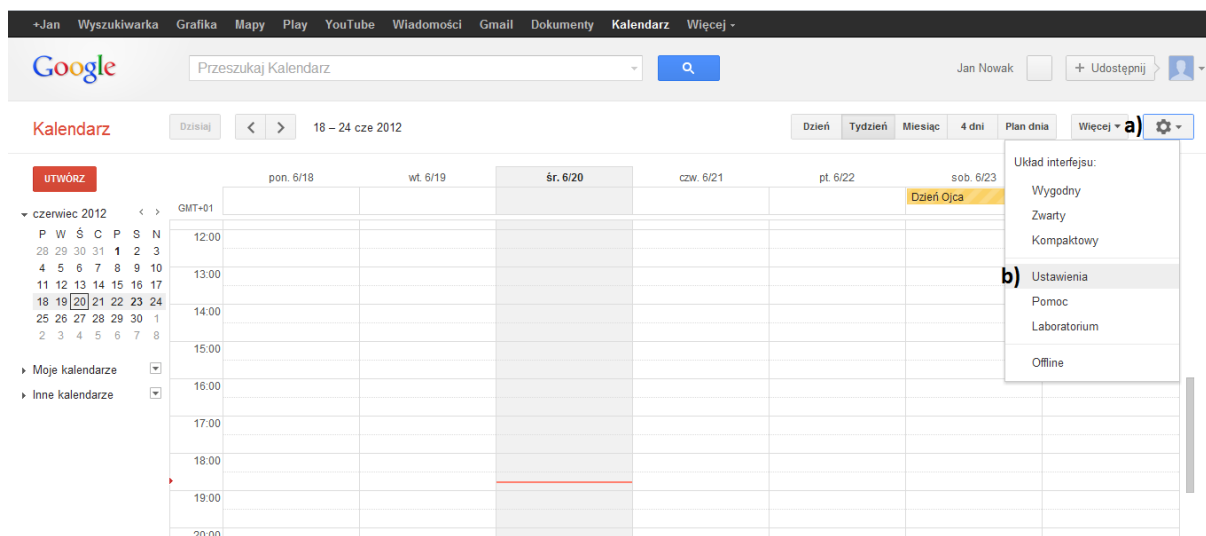
## Przed rozpoczęciem

Aby zapewnić poprawne działanie programu potrzebne jest konto Gmail. Zalecane jest utworzenie nowego konta, aby nie zaśmiecać swojego kalendarza czy skrzynki nadawczej dodatkowymi wpisami.

Nazwę użytkownika oraz hasło do konta należy podać w panelu ustawień użytkownika w zakładce Gmail.

Aby umożliwić poprawne wysyłanie wiadomości SMS należy dodatkowo aktywować usługę powiadomień smsowych na koncie Google. Można tego dokonać w następujący sposób:

### 1) Należy uruchomić ustawienia kalendarza Google:



Rys 5. Ustawienie powiadomień sms cz. 1.

- 2) Należy podać swój numer telefonu. Można to zrobić następująco:
  - a) Wejść w zakładkę ustawień telefonu komórkowego
  - b) Podać numer telefonu
  - c) Zatwierdzić
  - d) Przepisać kod weryfikacyjny otrzymany smsem na podany numer
  - e) Zakończyć konfigurację

Rys 6. Ustawienie powiadomień sms cz. 2.

## Szybkie wprowadzenie - przykładowy scenariusz

W podrozdziale zaprezentowany zostanie przykładowy scenariusz użycia polegający na utworzeniu nowej bazy przeszukiwań zawierającej 1 kanał RSS oraz 1 słowo kluczowe.

### Ustawienie danych użytkownika

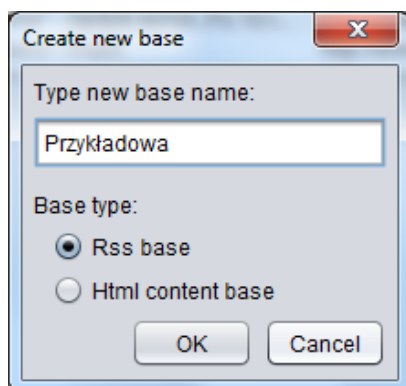
Pierwszym krokiem, jaki należy zrobić po pierwszym uruchomieniu programu jest ustawienie danych użytkownika poprzez naciśnięcie przycisku „User properties” w głównym oknie programu. Po uzupełnieniu danych, zamykamy okno.

Rys 7. Ustawienia danych użytkownika (przykład).



### Dodanie nowej bazy przeszukiwań

Następnie należy utworzyć bazę przeszukiwań, w której będą znajdować się kanały RSS. Można tego dokonać poprzez naciśnięcie „New base” w głównym oknie programu. Nazwijmy bazę „Przykładowa” oraz zaznaczmy jej typ jako baza kanałów RSS. Następnie zatwierdzamy przez „Ok”.



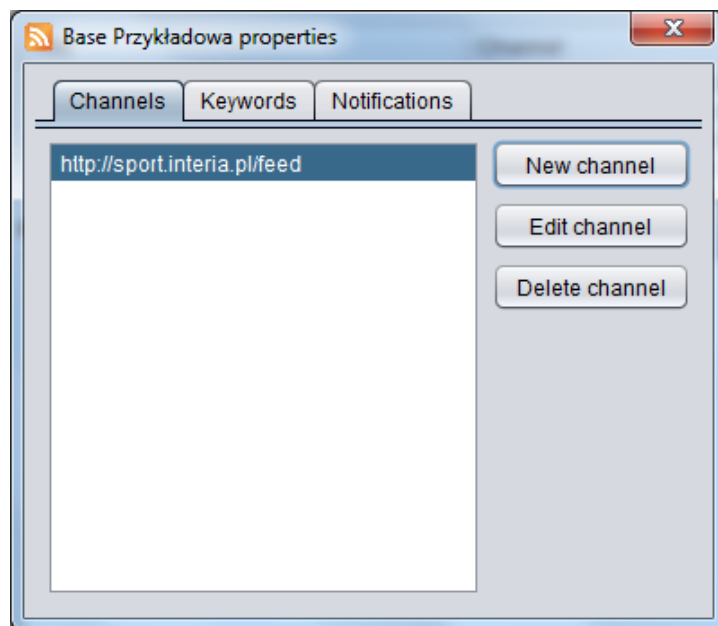
Rys 8. Tworzenie nowej bazy (przykład).

### Ustawienie parametrów bazy

Ostatnim krokiem jest ustawienie parametrów nowo utworzonej bazy. Należy zaznaczyć bazę na liście w głównym interfejsie a następnie kliknąć na „Base properties”. W oknie ustawień należy ustawić 3 zakładki:

### Dodanie nowego kanału RSS

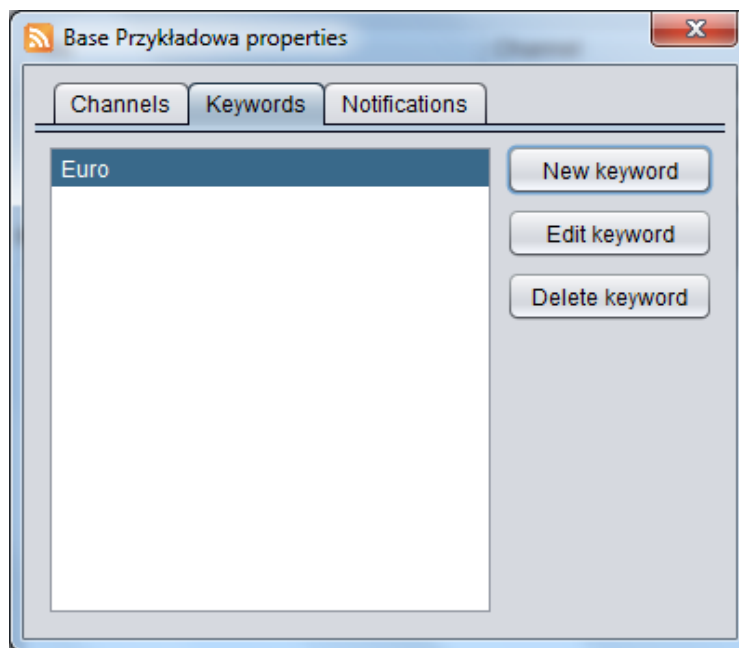
Poprzez naciśnięcie przycisku „New channel”



Rys 9. Ustawienia bazy – dodawanie kanału.

### *Dodanie nowego słowa kluczowego*

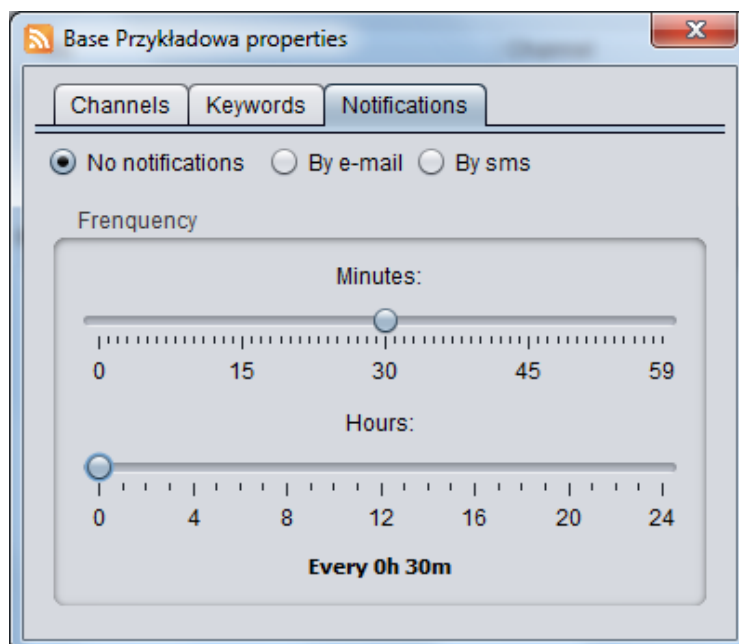
Poprzez naciśnięcie przycisku „New keyword”. Słowo kluczowe jest podstawą klasyfikowania kanałów przez program.



Rys 10. Ustawienia bazy – dodanie słowa kluczowego.

### *Ustawienie sposobu i częstotliwości powiadomień*

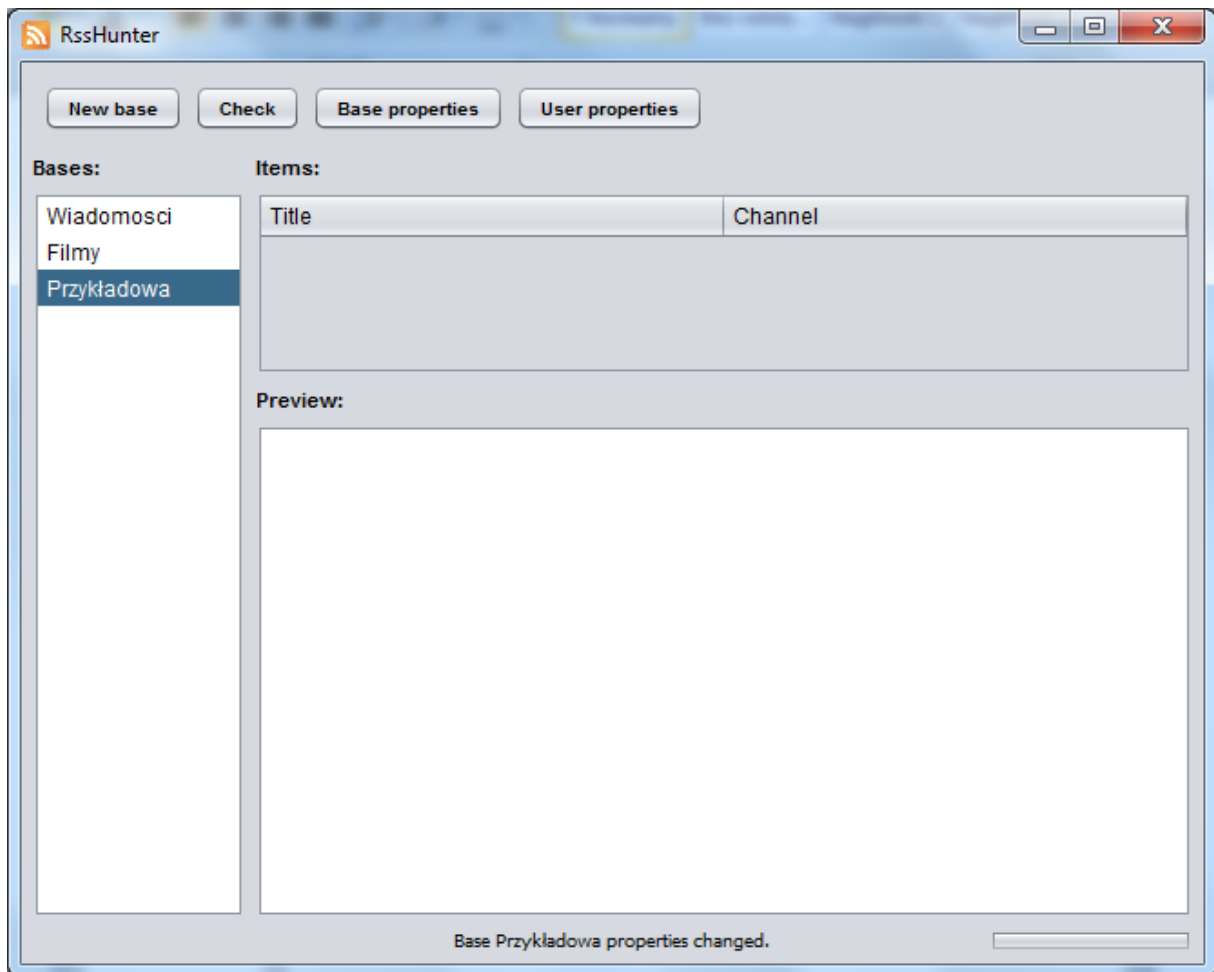
Jako, że Euro2012 jest popularnym tematem i nowe wpisy mogą pojawiać się często, ustawiamy brak dodatkowych powiadomień. Program będzie informował o nowych wpisach przez popup. Ustawiamy częstotliwość odświeżania bazy na 30min.



Rys 11. Ustawienie powiadomień (przykład).

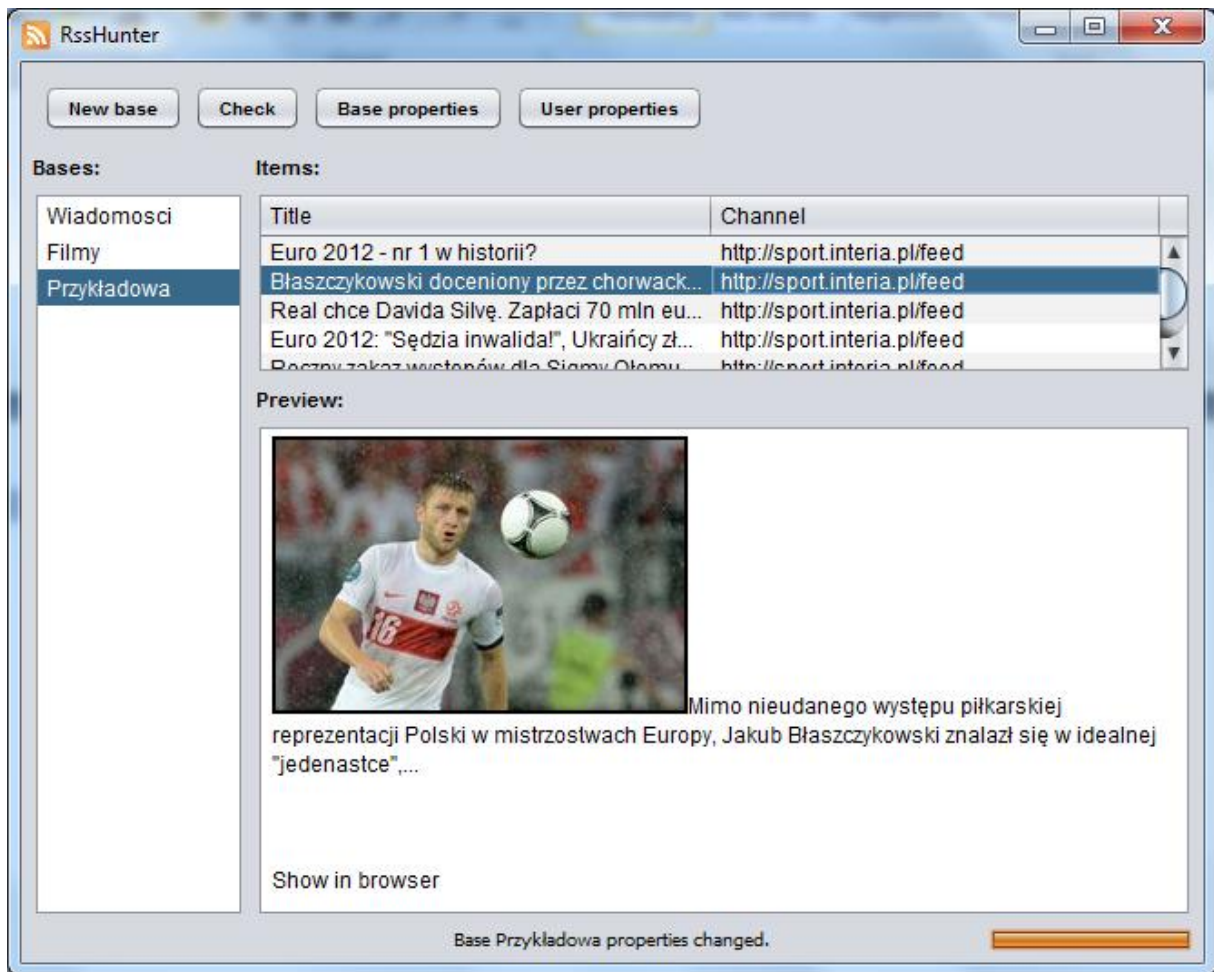
## Zakończenie

Baza została pomyślnie utworzona. Widać ją po lewej stronie na panelu baz przeszukiwań.



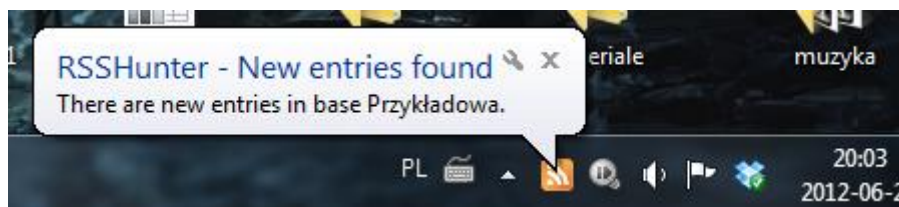
Rys 12. Nowa baza po dodaniu.

Możemy wymusić pobranie nowych wpisów przez kliknięcie na „Check”. Nowe wpisy można przeglądać wybierając je z listy.



Rys 13. Nowa baza po znalezieniu wpisów.

Przeczytanie całego wpisu możliwe jest przez kliknięcie w „Show In browser” na dole podglądu. Spowoduje to otwarcie pełnego artykułu w przeglądarce internetowej. Kiedy RssHunter jest schowany do ikony w trayu, a pojawią się nowe wpisy, użytkownik zostanie powiadomiony przez popup.



Rys 14. Powiadomienie popup przy nowych wpisach.

# Szczegóły implementacji

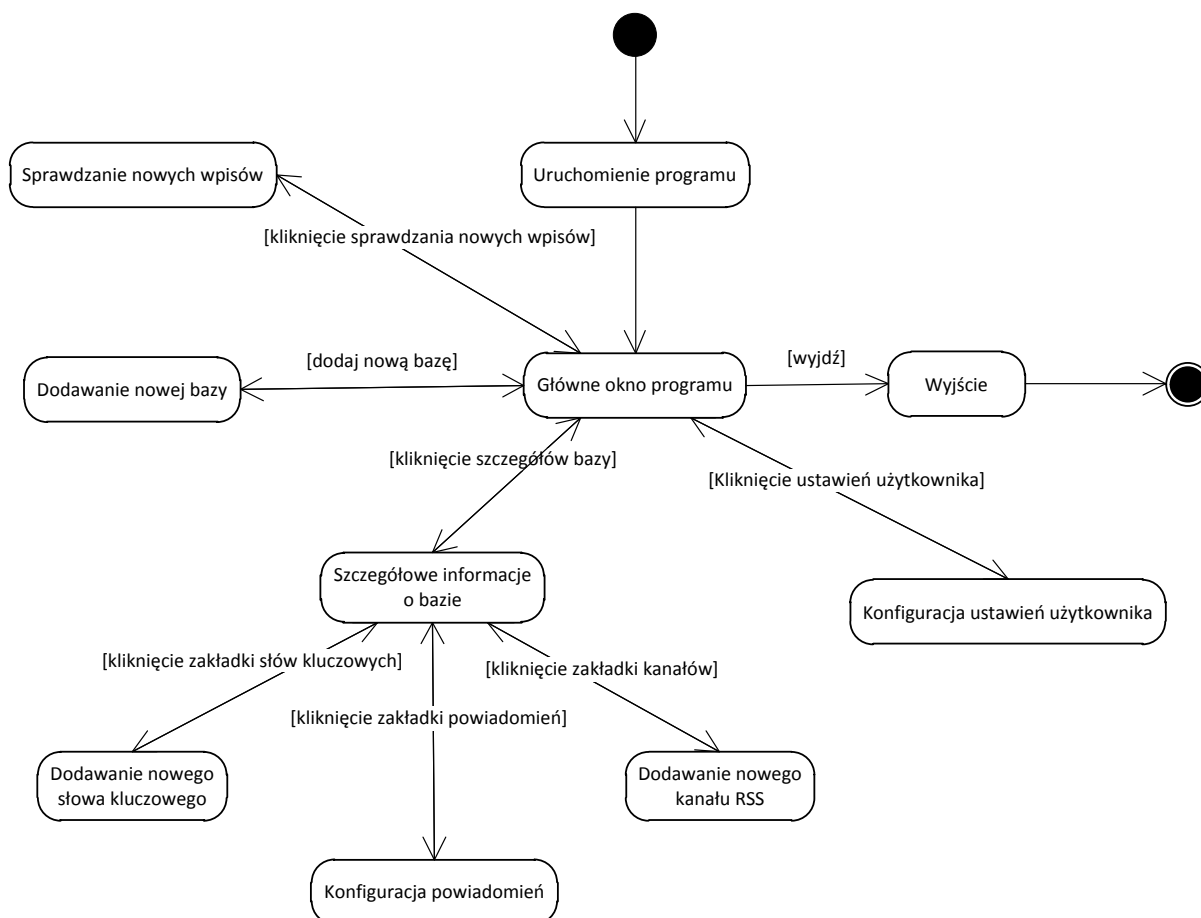
## Wykorzystane technologie i biblioteki

Logika programu została zaimplementowana w języku **Java SE6**, z wykorzystaniem biblioteki **Swing** do utworzenia interfejsu graficznego. Dodatkowe technologie wykorzystane w programie to:

- **JDBC** – do przechowywania danych w bazie **SQLite3**.
- **RSSHorrer** – parsowanie RSS
- **JavaMail** – powiadomienia mailowe
- **Google API** – powiadomienia SMSowe

## Diagram stanów

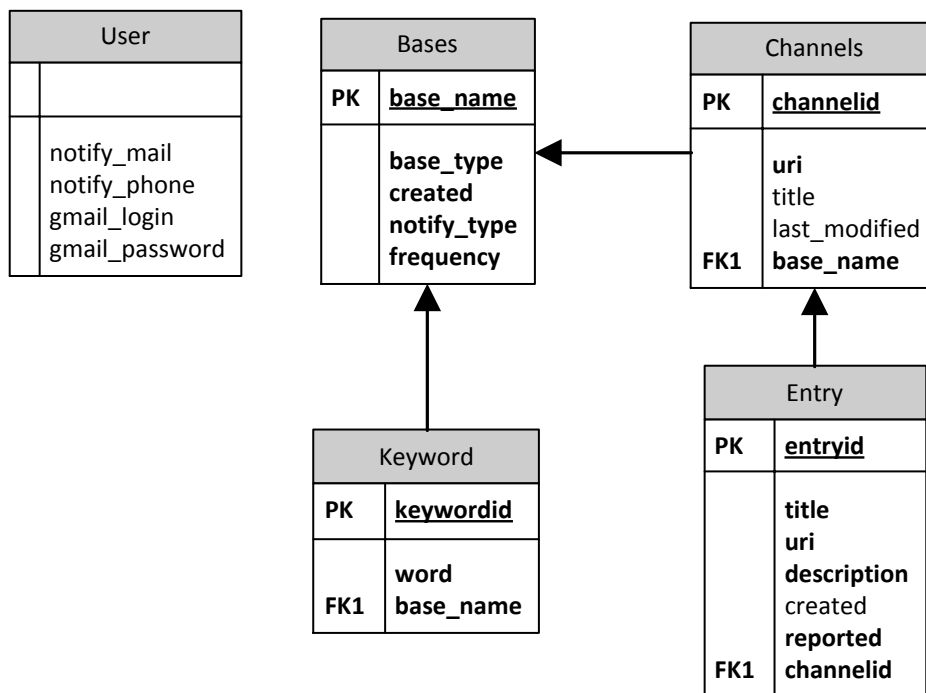
Diagram stanów programu prezentuje się następująco



Rys 15. Diagram stanów.

## Baza danych

Baza danych wykorzystana w programie ma następującą strukturę.



Rys 16. Schemat bazy danych

SQLite3 przechowuje dane bazy w pojedynczym pliku. W przypadku RssHuntera jest to plik data.db w głównym katalogu programu.

## Struktura programu

Program został podzielony na 4 paczki: Gui, Notifiers, Parsers, Rsshunter. Każda z nich jest odpowiedzialna za inną część logiki programu:

- GUI – zawiera klasy reprezentujące interfejs programu.
- Notifiers – klasy odpowiedzialne za powiadamianie użytkownika o nowych wpisach za pośrednictwem Sms/Mail.

- Parsers – abstrakcyjna klasa AbstractParser do tworzenia nowych parserów oraz klasy RssParser i HtmlParser do sprawdzania nowych wpisów w bazach typu RSS i Html.
- RssHunter – główna logika programu, klasa DB odpowiedzialna za kontakt z bazą danych, Encoder odpowiedzialna za szyfrowanie, klasy reprezentujące encje w bazie danych oraz klasa Scheduler odpowiedzialna za sprawdzanie baz przeszukiwań o odpowiedniej porze.

## Szczegóły działania programu

### Tworzenie baz przeszukiwań

Bazy przeszukiwań i wszystkie dane programu są przechowywane w pliku data.db w głównym katalogu programu. Po uruchomieniu programu, nazwy baz są ładowane do pamięci i przechowywane w modelu listy baz w kontenerze. Szczegółowe dane baz są ładowane w locie z bazy SQLite. Założeniem programu jest lekkość i jak najmniejsze obciążenie pamięci RAM (działa on głównie w tle i powiadamia użytkownika o zmianach, powinien być nieodczuwalny).

### Odświeżanie bazy RSS

Podczas odświeżania bazy RSS program najpierw sprawdza czy na kanale są nowe wpisy – obiekt reprezentujący kanał posiada pole lastModified zawierające datę ostatniej aktualizacji kanału RSS, z którego zostały pobrane wpisy. Program porównuje tą datę z aktualną datą lastModified kanału RSS. Jeśli się różnią pobierane są nowe wpisy. Następnie wpisy te są parsowane przy pomocy biblioteki RssHorror i sprawdzane pod kątem słów kluczowych wykorzystując klasy Pattern i Matcher w API JavySE. Jeśli jakieś wpisy pasują do słów kluczowych, są zapisywane w bazie. A użytkownik powiadamiany o nich w odpowiedni sposób.

### Odświeżanie bazy Html

Celem baz przeszukiwań html jest sprawdzanie czy zawartość strony się zmieniła. Taka zawartość musi być przechowywana. Rozwiązane jest to w następujący sposób: adresy stron html przechowywane są w bazie w taki sam sposób jak kanały RSS. Jednak po dodaniu nowej strony do bazy, natychmiastowo tworzony jest plik z jej zawartością w folderze HtmlContent. Nazwa pliku to id kanału, dzięki czemu pliki mogą być łatwo odnajdywane. Przy odświeżaniu bazy html pobierana jest aktualna zawartość strony i porównywana z tą w pliku. Użytkownik może wybrać tagi html, jakie mają być sprawdzane. W takim wypadku w pliku zapisywane są same zawartości wybranych tagów.

### Planowanie automatycznego odświeżania baz

Automatyczne odświeżanie baz odbywa się w klasie Scheduler. Klasa ta posiada wątek robotnika, który uruchamia odświeżanie poszczególnych baz. Poza tym tworzy też osobne wątki dla każdej bazy, które odliczają określony, ustawiony przez użytkownika czas. Po upływie czasu wątek odliczający powiadamia wątek odświeżania, aby sprawdził daną bazę przeszukiwań.

### Wysyłanie powiadomień mailowych

Powiadomienia mailowe są tworzone przez klasę MailBuilder na podstawie danych o nowych wpisach. Klasa tworzy wiadomość zawierającą powiadomienie o wystąpieniu nowych wpisów oraz raport dotyczący:

- W której bazie są nowe wpisy
- Listę kanałów, w których znaleziono nowe wpisy
- Ilość nowych wpisów

W przypadku bazy Html wysyłane jest jedynie powiadomienie o zmianie zawartości strony. Gdy wiadomość została już skomponowana, jest ona wysyłana przez klasę MailSender wykorzystującą bibliotekę JavaMail.

### Wysyłanie powiadomień SMS

Wysyłaniem SMSów zajmuje się klasa SmsSender wykorzystująca Google API. Wiadomości tego typu powinny być z założenia krótkie, więc użytkownik jest jedynie powiadamiany o zmianie zawartości strony lub o ilości nowych wpisów RSS w danej bazie. Powiadomienia SMS są darmowe, kosztem jest jednak wysyłanie ich przez Google Calendar. Gdy program chce wysłać SMS, tworzy nowe wydarzenie w kalendarzu użytkownika za 6 minut od aktualnego czasu. Nazwą wydarzenia jest treść powiadomienia. Następnie ustawia przypomnienie o wydarzeniu na 5min przed jego datą. W ten oto sposób po minucie użytkownik dostaje nową wiadomość sms przypominającą o wydarzeniu, którego nazwą jest treść od RssHuntera. Wiadomość taka zawiera dodatkowy opis od Google Calendar, jednak jej wysłanie nic nie kosztuje i spełnia swoje zadanie – powiadomienie użytkownika o nowych wpisach.