

The logo for 'PublMed Crawler 1.0' features the word 'PublMed' in a large, blue, sans-serif font. The 'M' is stylized with a white 'M' inside a blue shape that resembles a book or a stylized letter. Below 'PublMed' is the word 'Crawler 1.0' in a smaller, blue, sans-serif font.

PublMed Crawler 1.0

Projekt zaliczeniowy z przedmiotu *Informatyka w medycynie* - Politechnika Poznańska 2015 / 2016.

Autor: Sebastian Boruta

Prowadzący: dr hab. inż. Szymon Wilk

Spis treści

1. Dokumentacja użytkowa

- 1.1. O aplikacji
- 1.2. Pierwsze uruchomienie
- 1.3. Przedstawienie interfejsu
 - 1.3.1. Ekran startowy
 - 1.3.2. Pomoc
 - 1.3.3. Aktualizacja lokalnej bazy
 - 1.3.4. Przeszukiwanie lokalnej bazy
 - 1.3.5. Podgląd dokumentu

2. Dokumentacja deweloperska

- 2.1. Wykorzystane technologie i narzędzia
- 2.2. Architektura systemu
- 2.3. Uruchomienie projektu w środowisku NetBeans
- 2.4. Przegląd klas i pakietów

1. Dokumentacja użytkowa

1.1. O aplikacji

Aplikacja PubMed Crawler 1.0 została stworzona na potrzeby zaliczenia z przedmiotu „Informatyka w medycynie” prowadzonego przez dr hab. inż. Szymona Wilka na Politechnice Poznańskiej.

Głównym zadaniem programu jest utworzenie lokalnej bazy danych publikacji medycznych w oparciu o bazę PubMed dla wybranych kategorii. Dodatkowo aplikacja daje użytkownikowi możliwość przeszukiwania lokalnej kopii danych z wykorzystaniem silnika Lucene, oraz przeglądania wyników z możliwością łatwego odnalezienia właściwych wpisów w serwisie PubMed.

Użytkownikami systemu mogą być np. lekarze, którzy prowadzą badania na wybrany temat i poszukują artykułów dotyczących danej tematyki. Przy pomocy tej aplikacji będą oni mogli w łatwy sposób pobrać i lokalnie przeszukiwać interesującą ich część bazy PubMed, oraz w łatwy sposób ją aktualizować.

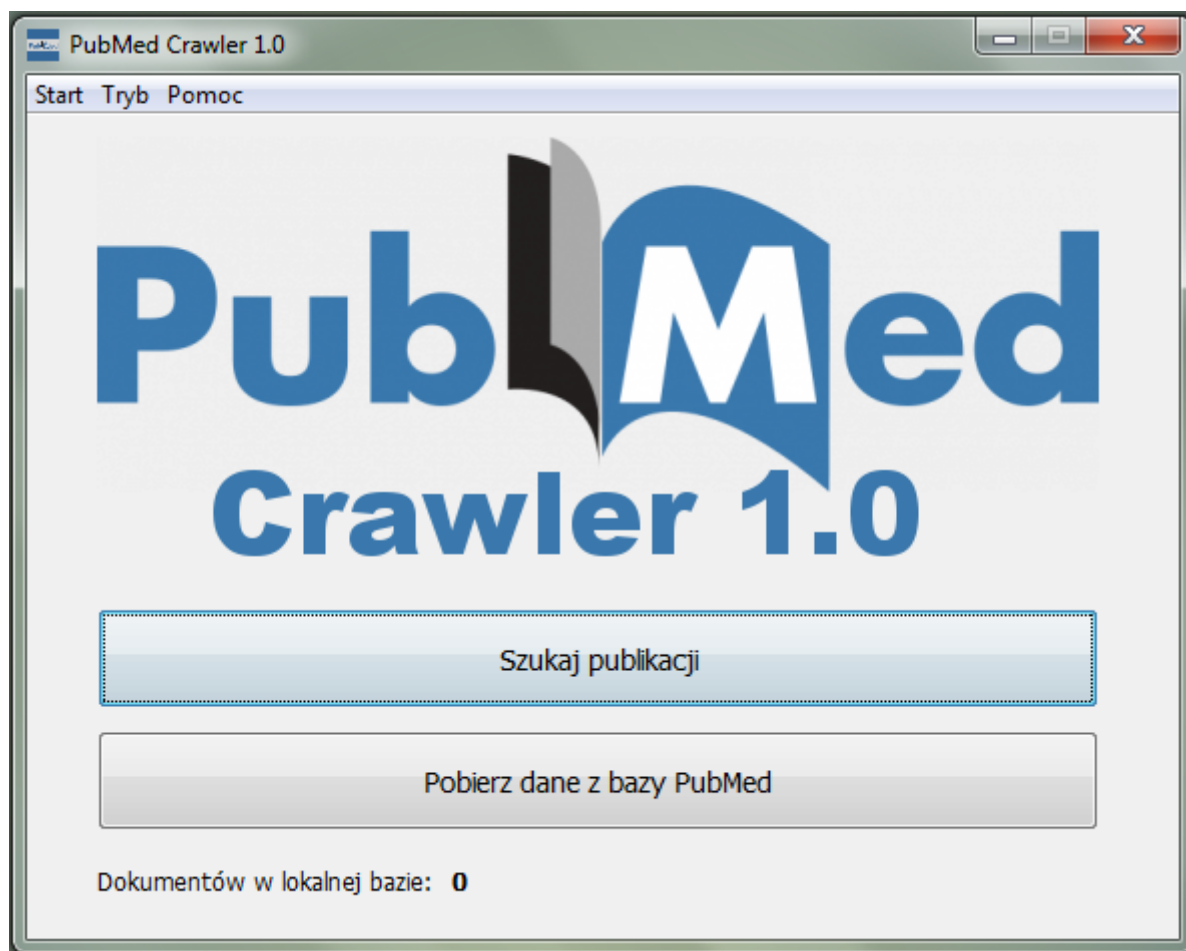
1.2. Pierwsze uruchomienie

Stabilna wersja aplikacji znajduje się w archiwum *PubMed Crawler 1.0 (release).rar* w głównym katalogu repozytorium. Aby z niej skorzystać, wystarczy rozpakować archiwum i uruchomić plik *PubMed_crawler.jar*. Przy pierwszym uruchomieniu stworzony zostanie dodatkowy folder na indeks Lucene, oraz plik *settings.db* w którym znajdują się ustawienia aplikacji.

1.3. Przedstawienie interfejsu

1.3.1. Ekran startowy

Jest to pierwszy widok po uruchomieniu aplikacji. Można również uruchomić go po wyborze opcji *Pokaż stronę startową* z menu *Start*.



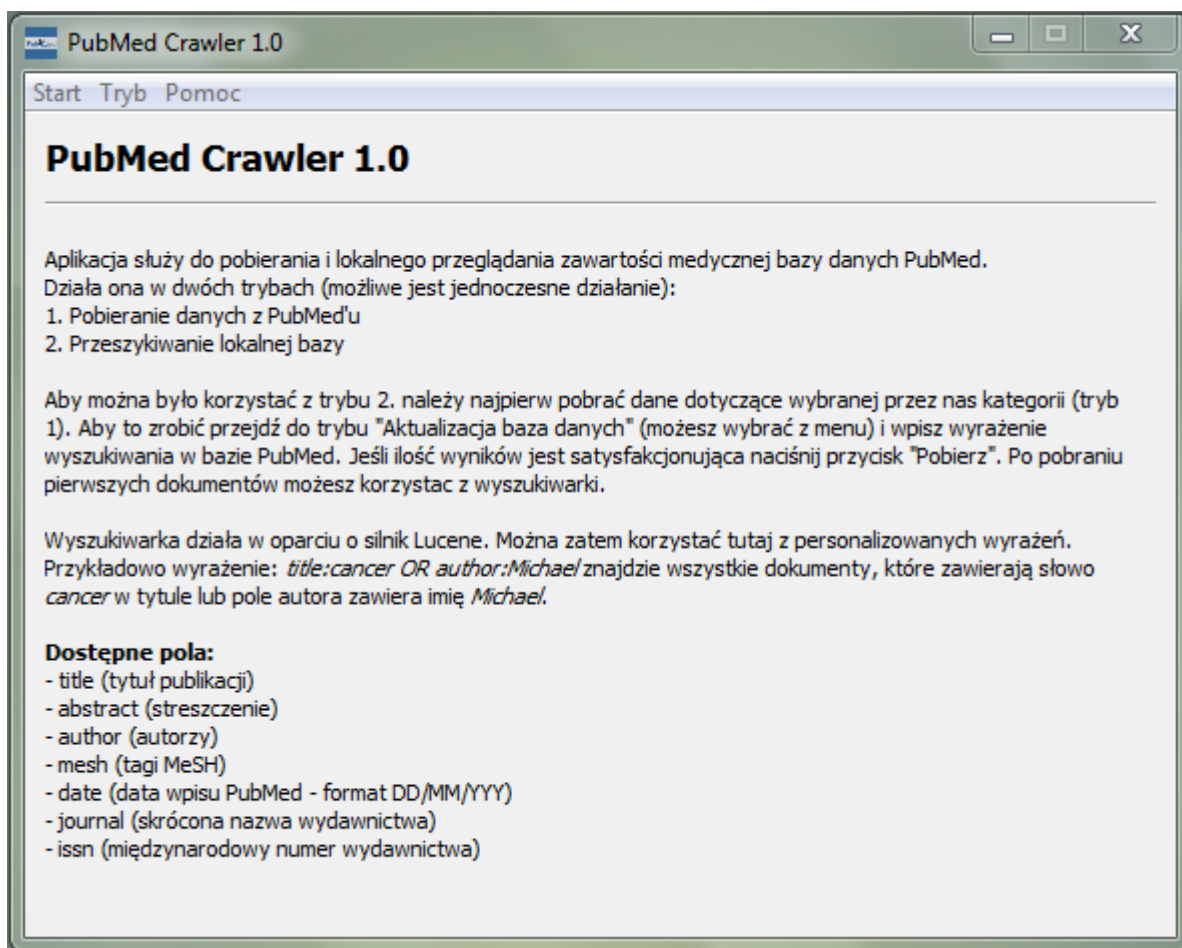
Poza wielkim logo znajdują się tutaj dwa przyciski, które służą do szybkiej nawigacji po głównych funkcjach systemu:

- *Szukaj publikacji* przenosi nas do trybu wyszukiwania publikacji
- *Pobierz dane z bazy PubMed* przenosi nas do trybu pobierania danych

Dodatkowo w tym widoku znajduje się informacja o sumarycznej liczbie publikacji, o których informacje znajdują się w lokalnej bazie danych.

1.3.2. Pomoc

W tej części aplikacji znajdują się podstawowe informacje na temat jej działania, oraz przydatne informacje dotyczące wyszukiwania przy użyciu silnika Lucene. Aby ją uruchomić należy wybrać opcję *Pomoc* z menu *Pomoc*.



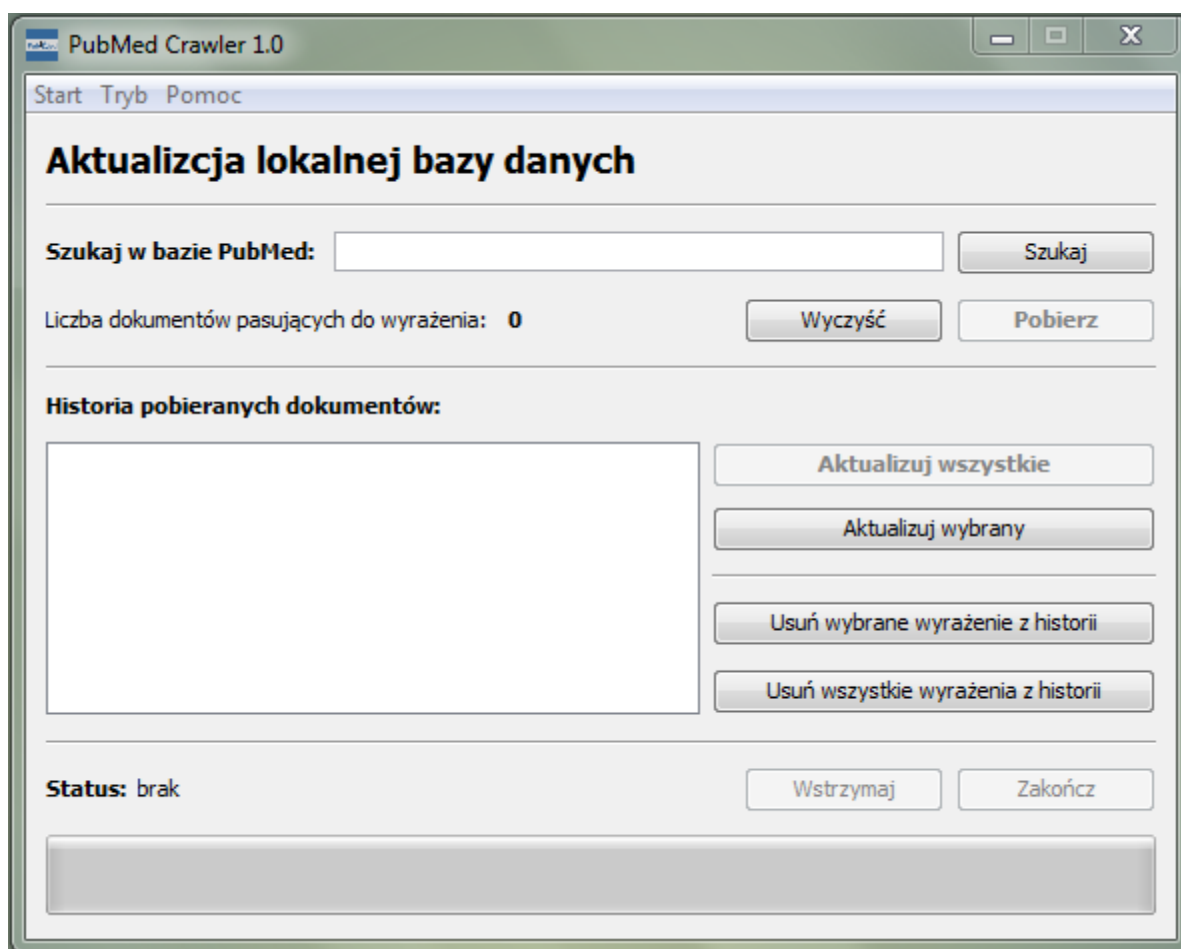
Nie ma tutaj żadnych dodatkowych przycisków. Warto zapoznać się z treścią tej strony, ponieważ znajduje się tutaj wiele przydatnych informacji, pomagających w efektywny sposób korzystać z aplikacji.

W celu pogłębienia wiedzy dotyczącej wyszukiwania przy pomocy Lucene, warto zapoznać się z innymi źródłami dostępnymi za darmo w internecie. Przykładowe strony po polsku:

- http://weblis.ibles.pl/libcat/help/Lucene_query_language.html
- <http://jdn.pl/node/33>

1.3.3. Aktualizacja lokalnej bazy danych

Jest to jedna z dwóch głównych sekcji aplikacji. Pozwala ona na pobieranie danych z bazy PubMed. Aby przejść do tego widoku należy wybrać *Aktualizacja bazy danych* z menu *Tryb* lub kliknąć w przycisk *Pobierz dane z bazy PubMed* w ekranie startowym aplikacji.



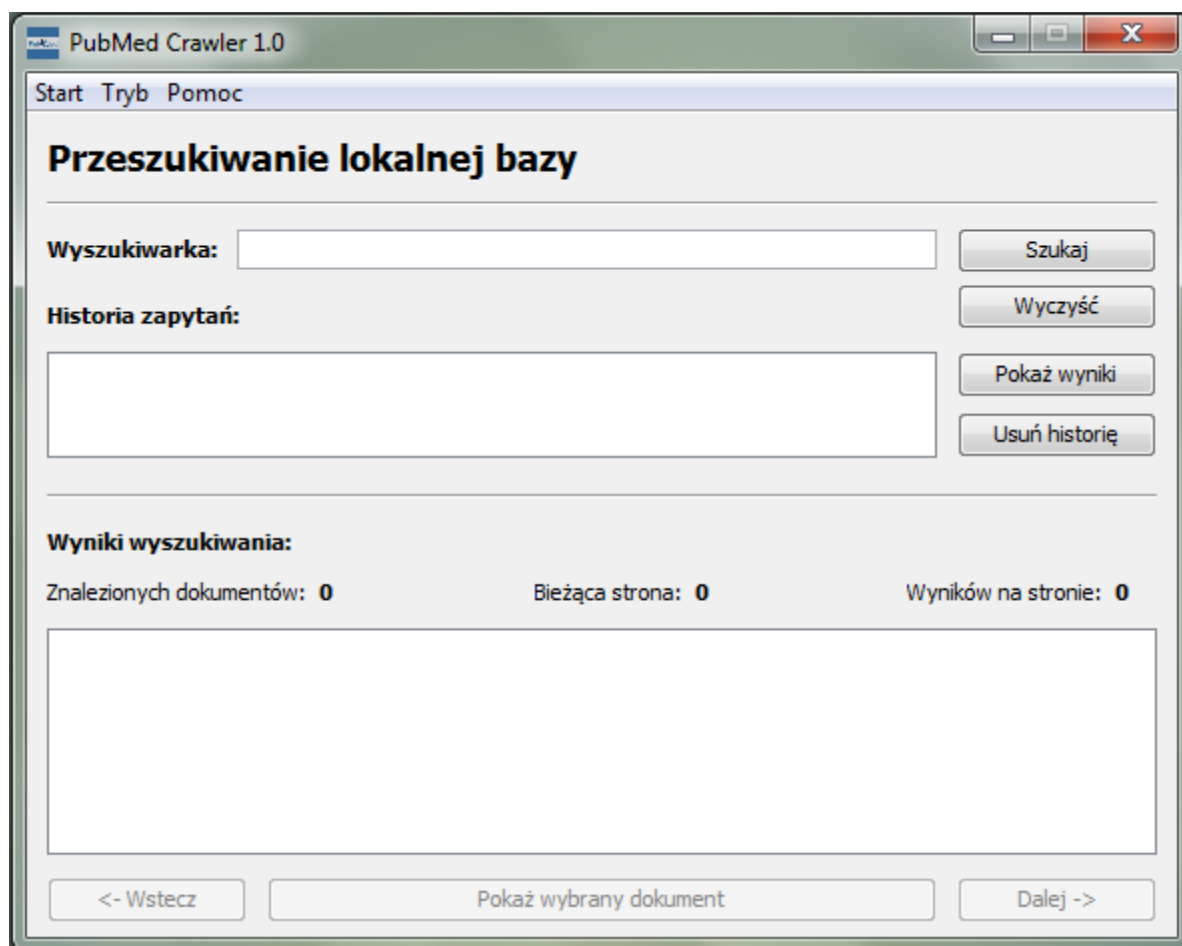
W celu odnalezienia publikacji wpisujemy w pole tekstowe wyrażenie wyszukiwania PubMed (dowolne wyrażenie, zgodne z wyszukiwarką PubMed), a następnie klikamy przycisk *Szukaj*. Aplikacja wyśle zapytanie do serwisu i zwróci nam informację z liczbą wyników.

Jeżeli jesteśmy zainteresowani pobraniem danych, należy wcisnąć przycisk *Pobierz*. W przeciwnym wypadku klikamy *Wyczyść*. Pobieranie danych można wstrzymać lub przerwać w dowolnym momencie (przyciski *Wstrzymaj* oraz *Zakończ*).

Dodatkowo znajduje się tutaj historia, dzięki której możemy łatwo aktualizować nasz spis publikacji. Wyczyszczenie historii nie powoduje usunięcia danych z bazy.

1.3.4. Przeszukiwanie lokalnej bazy

Jest to druga główna sekcja aplikacji. Pozwala ona na wyszukiwanie publikacji w lokalnej bazie danych. Przejście do tego widoku umożliwia nam przycisk *Szukaj publikacji* dostępny na ekranie startowym, lub wybór *Przeszukiwanie lokalnej bazy* z menu *Tryb*.



The screenshot shows a window titled "PubMed Crawler 1.0" with a menu bar containing "Start", "Tryb", and "Pomoc". The main content area is titled "Przeszukiwanie lokalnej bazy". It features a search input field labeled "Wyszukiwarka:" with a "Szukaj" button to its right. Below this is a section labeled "Historia zapytań:" with a large empty text area and buttons "Wyczyść", "Pokaż wyniki", and "Usuń historię" to its right. At the bottom of the main area, under "Wyniki wyszukiwania:", it displays "Znalezionych dokumentów: 0", "Bieżąca strona: 0", and "Wyników na stronie: 0". A large empty box is provided for the search results. At the very bottom of the window are three navigation buttons: "<- Wstecz", "Pokaż wybrany dokument", and "Dalej ->".

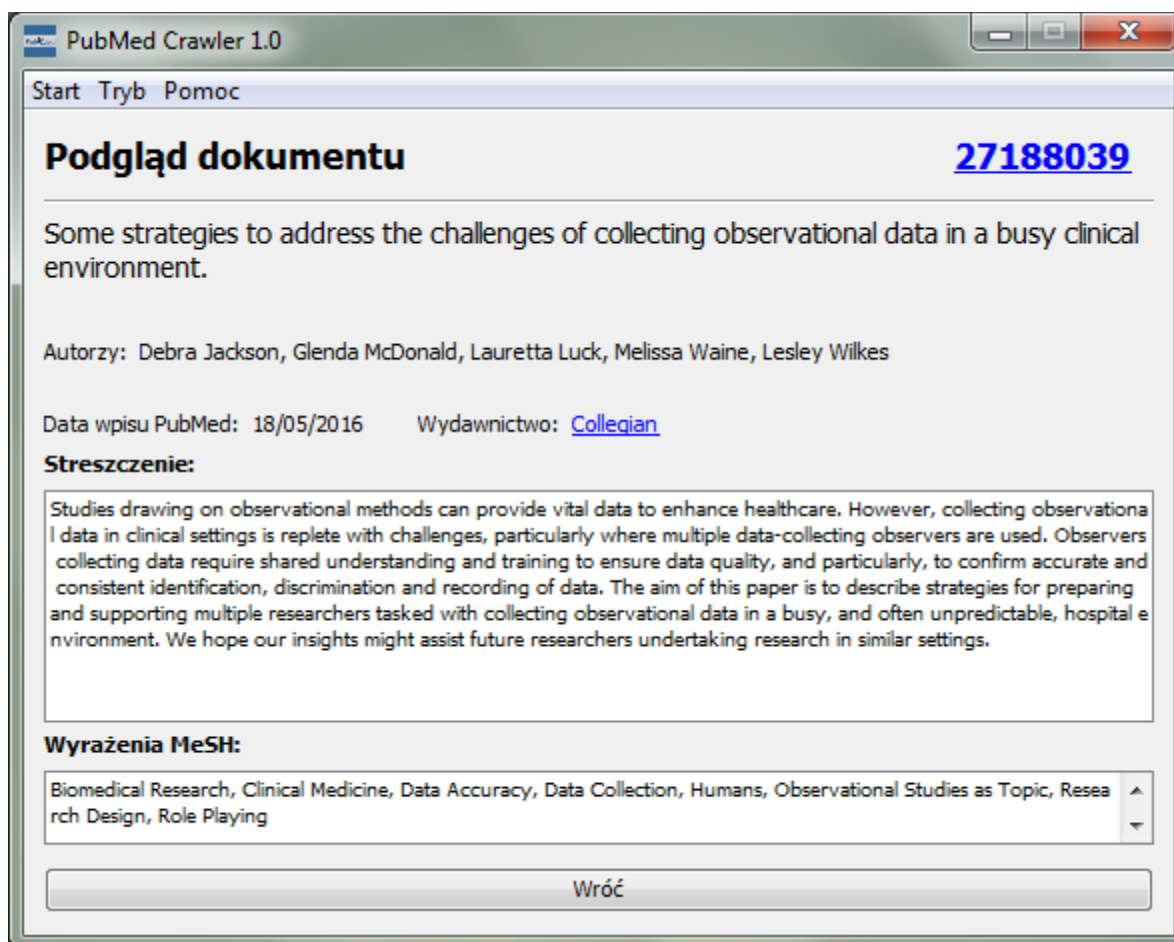
Aby wyszukać interesujące nas dokumenty, należy wypełnić pole wyszukiwarki (można posłużyć się tutaj wyrażeniami Lucene), a następnie kliknąć przycisk *Szukaj*. W zależności od liczby pasujących publikacji, wyniki pojawią się w dolnej sekcji w ilości po 5 na stronę. Nawigację pomiędzy stronami umożliwiają przyciski *Wstecz* oraz *Dalej*.

Aby przejść do konkretnego dokumentu, należy wybrać go z listy, a następnie kliknąć na środkowy przycisk *Pokaż wybrany dokument*.

Wszystkie wyszukiwane wyrażenia zapisywane są w historii, którą można łatwo usunąć, lub pokazać wyniki dla wybranego, archiwalnego wyrażenia.

1.3.5. Podgląd dokumentu

Ostatni z widoków dostępny jest po wybraniu publikacji z listy wyników wyszukiwarki (patrz punkt wyżej). Znajdują się tutaj szczegółowe informacje dotyczące wybranego dokumentu.



W prawym górnym rogu znajduje się identyfikator publikacji w bazie PubMed – po kliknięciu w niego zostaniemy przeniesieni do oryginalnego wpisu na w.w. stronie.

Nieco niżej znajduje się tytuł dokumentu, oraz informacje o autorach, dacie wpisu oraz wydawnictwie. Nazwa wydawnictwa jest odnośnikiem prowadzącym do wpisu na stronie [issn.cc](http://www.issn.cc), zawierającej dodatkowe dane wydawcy.

W białych polach tekstowych znajduje się streszczenie artykułu oraz lista wyrażeń MeSH użytych do klasyfikacji tematyki artykułu.

2. Dokumentacja deweloperska

2.1. Wykorzystane narzędzia i technologie

- Język programowania: Java
- Biblioteka graficzna (GUI): Swing
- Przechowywanie danych: SQLite 3+, Lucene
- Przetwarzanie danych: SAX
- Środowisko programistyczne: NetBeans IDE 8.1
- System operacyjny: Windows 7 x64

2.2. Architektura systemu

System składa się z kilku modułów, które wspólnie tworzą funkcjonalną aplikację:

- interfejs komunikacji z użytkownikiem (GUI)
- moduł komunikacji sieciowej
- moduł komunikacji z lokalną bazą SQLite
- moduł parsowania wyników z PubMed
- moduł odpowiedzialny z pracę z Lucene
- moduł zarządzający pracą wielu wątków

2.3. Uruchomienie projektu w środowisku NetBeans

Uruchomienie projektu jest bardzo proste. Dla ułatwienia w repozytorium załączone są wykorzystywane biblioteki, zatem nie trzeba ich dodatkowo pobierać.

1. Pobieramy skompresowane repozytorium z GitHub'a
2. Wypakowujemy plik w wybranym folderze
3. Otwieramy środowisko programistyczne NetBeans
4. Wybieramy *File -> Open Project*
5. Przechodzimy do katalogu, w którym rozpakowaliśmy archiwum, a następnie wybieramy katalog *JavaCrawler* i klikamy przycisk *Open Project*
6. Na zakładce *Projects* aplikacji NetBeans sprawdzamy czy w folderze *Libraries* znajdują się niezbędne biblioteki (1 x *sqlite*, 3 x *lucene*). Jeśli nie zostały dodane automatycznie, można to zrobić ręcznie we właściwościach projektu (*Properties -> Libraries*).

Pod wykonaniu powyższych kroków aplikacja powinna działać poprawnie (można to sprawdzić uruchamiając projekt). Można zacząć pracę z projektem. Jeżeli po wykonaniu wszystkich kroków aplikacja nie chce się zbudować spróbuj użyć starszej wersji środowiska NetBeans i / lub JDK.

2.4. Przegląd klas i pakietów

Aplikacja składa się z 7 pakietów (ang. packages), z których każdy zawiera co najmniej jedną klasę. Poniżej znajdują się ich lista oraz krótki opis ich wykorzystania:

- **Main** – podstawowe działanie
 - *Crawler.java* – główna klasa, która uruchamia GUI
- **GUI** – wygląd aplikacji
 - *CrawlerGUI.java* – klasa odpowiedzialna za działanie GUI
 - *Icon.png*, *logo.png* – pliki graficzne wykorzystywane w aplikacji

- **SQLite** – obsługa bazy danych
 - *DB.java* – klasa odpowiedzialna za obsługę bazy SQLite
- **Lucene** – obsługa silnika Lucene
 - *Lucene.java* – klasa odpowiedzialna za obsługę silnika Lucene
- **ParserXML** – parsowanie dokumentów XML
 - *PubMed.java* – główna klasa parsująca
 - *SearchHandler.java* – klasa parsująca wyniki wyszukiwania
 - *FetchHandler.java* – klasa parsująca szczegółowe wpisy PM
 - *SummaryHandler.java* – klasa parsująca skrócone wpisy PM
- **Structs** – struktury wykorzystywane przez programistę
 - *Summary.java* – klasa, której obiekty reprezentują wpisy PM
 - *Results.java* – klasa, której obiekty reprezentują wyniki wyszukiwania
- **Other** – inne elementy aplikacji
 - *Task.java* – klasa, która pierwotnie miała służyć do obsługi wątków

Kolorem szarym zaznaczone zostały pliki, które w obecnej wersji nie są wykorzystywane do prawidłowego działania aplikacji.