Uniwersytet Mikołaja Kopernika
w Toruniu

Wydział Matematyki i Informatyki

Wydział Fizyki, Astronomii
i Informatyki Stosowanej

Robert Ospara
Nr albumu: 146951

Praca magisterska

na kierunku informatyka

Wizualizacja dopasowania wyrażeń regularnych Perla z użyciem biblioteki Gtk+

Praca wykonana pod kierunkiem
dr hab. Jacka Kobusa
Zakład Mechaniki Kwantowej

TORUŃ 2009

Spis treści

Wstęp

Wyrażenia regularne, potocznie zwane regexami (z ang. *regular expression*), pozwalają na zwięzły i elastyczny opis zbiorów słów. Mechanizm wyrażeń regularnych jest wykorzystywany przez szereg popularnych języków programowania (Perl, Java, Ruby, Python), edytory tekstów (sed, grep, find, vi, emacs), systemy bazodanowe, itp. Dla każdego wyrażenia regularnego można zbudować odpowiadający mu automat skończony. Automaty dzielimy na deterministyczne automaty skończone (DFA, z ang. *Deterministic Finite Automata*) oraz niedeterministyczne automaty skończone (NFA, z ang. *Nondeterministic Finite Automata*). Mechanizm DFA działa szybciej, lecz w przeciwieństwie do mechanizmu NFA nie pozwala na implementację takich cech jak przechwytywanie, przewidywanie, bądź też użycie kwantyfikatorów niezachłannych. Z kolei mechanizmy NFA można podzielić na dwie podklasy: tradycyjne oraz POSIX-owe [1]. W praktyce spotyka się również łączenie obu mechanizmów w zależności od wyrażenia i kontekstu, aby wykorzystać najlepsze cechy obu z tych mechanizmów. Jeden z najprostszych pomysłów polega na przeszukaniu łańcucha silnikiem DFA w celu sprawdzenia, czy dopasowanie w ogóle się powiedzie, a następnie wykorzystanie wolniejszego mechanizmu NFA dla określenia wyniku dopasowania. W języku interpretowanym Perl wyrażenia regularne są implementowane w oparciu o mechanizm niedeterministycznych automatów skończonych.

T.Kojm podjął próbę stworzenie narzędzia pozwalającego na obserwację działania mechanizmu wyrażeń regularnych w procesie dopasowania wzorca w Perlu [2]. Tak powstał program revis (ang. *Regular Expression VISualiser*) napisany w Perlu, którego działanie opiera się na analizie danych dostarczonych przez interpreter Perla uruchomiony w trybie debuggowania wyrażeń regularnych. Program ten dostarcza tekstowego interfejsu użytkownika dzięki wykorzystaniu modułu Curses. Zasadniczym celem niniejszej pracy jest rozwinięcie tego pomysłu przez zbudowanie interfejsu graficznego w oparciu o bibliotekę GTK w wersji 2 dla Perla 5.8.x (program grevis), ulepszenie wizualizacji procesu dopasowania, poszerzenie możliwości programu revis oraz usunięcie jego ograniczeń i błędów. Program grevis może stanowić pomoc dla wszystkich próbujących nauczyć się stosować wyrażenia regularne w Perlu. Otóż zwykle programiści piszący skrypty perlowe muszą testować używane wyrażenia regularne w oddzielnych programach. Często nie rozumieją oni do końca jak przebiega dopasowanie i dlatego skazani są na stosowanie metody prób i błędów. W takich sytuacjach pomocne może być użycie debuggera

wyrażeń regularnych. Jego działanie można zobaczyć uruchamiając w powłoce np. takie polecenie:

```
perl -e 'use re debug; \
        "aaaaaaaaaa" =~ /a?a?a?a?a?a?a?a?a?a?aaaaaaaaaa/'
```

Debugger dostarcza dużej ilości informacji, których analiza i zrozumienie nie jest łatwe. Zadaniem programu grevis jest przekształcenie tych danych do postaci wygodnej i zrozumiałej dla zwykłego użytkownika.

Istnieje szereg projektów, które pomagają w budowie i debuggowaniu wyrażeń regularnych. Warto wymienić następujące: KOMODO[1], REGEXBUDDY[2], regex-coach[3], regex editor[4], EditPad Pro[5], Expresso[6], rozszerzenia dla Microsoft Visual Studio[7], rozszerzenia dla Eclipsa[8]. W przeciwieństwie do tych programów grevis ukazuje krok po kroku procesu dopasowania w połączeniu ze zmianami w stanach automatu NFA.

Treść tej pracy zawiera opis wszystkich elementów GUI aplikacji grevis, przykłady użycia programu, wraz z analizą wyświetlanych informacji oraz dokumentację dotyczącą implementacji projektu. Do pracy została dołączona płyta CD-ROM, na której umieszczone zostały źródła pracy w formacie LaTeX oraz program grevis. Oprogramowanie zostało udostępnione na zasadach GNU General Public License.

[1] http://aspn.activestate.com
[2] http://www.regexbuddy.com
[3] http://weitz.de/regex-coach
[4] http://myregexp.com
[5] http://www.regular-expressions.info/editpadpro.html
[6] http://www.ultrapico.com/Expresso.htm
[7] http://www.microsoft.com/visualstudio/en-us/default.mspx
[8] http://www.eclipse.org