

Factorer MPI - dokumentacja projektowa

Rafał Duraj, Piotr Dowgiałło, Bartosz Janusz, Maciej Kolański

2016-06-07

Spis treści

1	Temat projektu	2
2	Zakres projektu	2
2.1	Cel	2
2.2	Funkcjonalność	2
3	Narzędzia i technologie zastosowane w projekcie	3
3.1	Zastosowane technologie	3
3.2	Narzędzia wykorzystane w projekcie	4
4	Aktualny stan rynku	4
5	Projekt techniczny	5
5.1	Klaster obliczeniowy	5
5.2	Strona internetowa	5
6	Dokumentacja powykonawcza (instalacyjna)	10
6.1	Uruchamianie strony	10
7	Istotne elementy kodu z komentarzem	11
8	Przykładowe wyniki badań efektywności programu równoległego	11
9	Wnioski	11

1 Temat projektu

W dzisiejszych czasach, gdy właściwie wszystko co robimy w jakiś sposób połączone jest z Internetem bezpieczeństwo jest bardzo ważnym tematem.

Faktoryzacja jest to proces podczas którego dla zadanego obiektu odnajduje się inne obiekty, które spełniają to, że ich iloczyn równy jest oryginalnemu obiektowi, w związku z czym te znalezione czynniki są w pewnym sensie od niego prostsze.

Podstawowy algorytm faktoryzacji bazuje na próbowaniu podziału liczby do faktoryzacji n przez wszystkie liczby pierwsze od 2 do \sqrt{n} . Tego typu algorytm bardzo dobrze radzi sobie z początkiem faktoryzacji liczby, bo dowolna liczba ma czynnik zarówno małe jak i duże. Jak wiadomo połowa wszystkich liczb dzieli się przez dwa, jedna trzecia liczb przez trzy i tak dalej, a więc z dużym prawdopodobieństwem można pozbyć się w prosty sposób niskich czynników.

RSA jest to jeden z pierwszych i też obecnie najpopularniejszych asymetrycznych algorytmów kryptograficznych gdzie klucz jest publiczny. Bezpieczeństwo szyfrowania przy pomocy tego algorytmu jest związane z trudnością faktoryzacji dużych liczb.

2 Zakres projektu

2.1 Cel

Celem projektu jest umożliwienie zlecenia zadania faktoryzacji dużej liczby (większych od $2^{64} - 1$). Jednym z głównych założeń jest udostępnienie prostego w obsłudze interfejsu użytkownika i zmaksymalizowanie elastyczności – system powinien być zdolny do współpracy z zadanymi komputerami, a instalacja wymaganego oprogramowania musi być prosta.

2.2 Funkcjonalność

Podstawowa

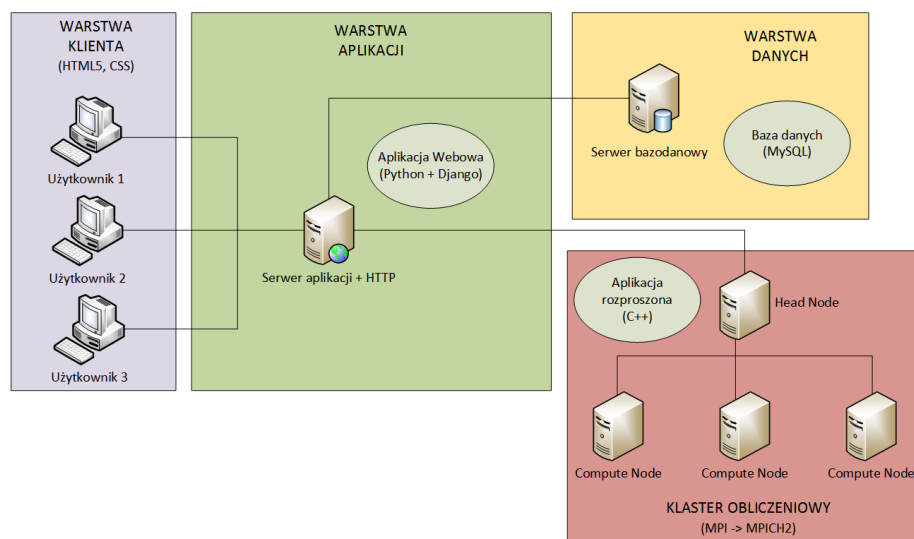
1. Udostępnienie mechanizmów rejestracji i logowania (konta użytkowników)
2. Zarządzanie zadaniami z poziomu interfejsu webowego
 - (a) zlecenie zadań
 - (b) przegląd historii
3. Możliwość rozbudowy klastra bez ingerencji w kod
4. Faktoryzacja metodą „brutalnej siły”
5. Faktoryzacja metodą CFRAC
6. Zachowywanie wyników pomyślanie wykonanych faktoryzacji

Rozszerzona

1. Łamanie szyfru RSA
2. Faktoryzacja metodą sita kwadratowego
3. Forma graficznej prezentacji wyników pomiarów

3 Narzędzia i technologie zastosowane w projekcie

3.1 Zastosowane technologie



Rysunek 1: Schemat blokowy systemu

Strona klienta - HTML5, CSS Interfejs z którego będzie korzystał klient projektowanego systemu został napisany w języku skryptowym HTML5 z użyciem stylów CSS. HTML5 jest obecnie standardem przy tworzeniu stron internetowych i w większości wyparł HTML4, w którego specyfikacji było dużo niescisłości. Użycie CSS z kolei pozwoli ujednolicić prezentację zawartości w różnych przeglądarkach, oraz uprości organizację samego kodu.

Aplikacja Webowa - Python 3.4 + Django 1.8.7[1] Python jest językiem wysokiego poziomu ogólnego przeznaczenia, z kolei Django to framework webowy dla tego języka. Wybraliśmy ten zestaw z powodu wielu ułatwień przy

tworzeniu aplikacji webowych, które są przezeń oferowane, np. dynamiczny interfejs bazy danych, automatyczny interfejs administracyjny. Dodatkowo część naszej grupy jest zaznajomiona z tymi technologiami, więc nie ma potrzeby poznawania ich od zera. Istotne są tutaj wersje środowisk - najnowsze dystrybucje nie obsługują MySQL, w związku z tym wybraliśmy poprzednie.

Baza danych - MySQL SZBD rozwijany przez firmę Oracle. Charakteryzuje się wszystkimi najważniejszymi funkcjonalnościami bazy danych oraz prostotą tworzenia takiej bazy. Rozważaliśmy zastosowanie systemu Oracle Database, jednakże jest on zbyt rozbudowany jak na nasze potrzeby, a co za tym idzie trudny w obsłudze. Mamy również doświadczenie w integracji bazy MySQL z aplikacjami napisanymi w Pythonie (Django).

Klaster obliczeniowy - standard MPI[2] MPICH2 to darmowa implementacja standardu MPI dla systemów Linux. Umożliwia proste tworzenie klastrów obliczeniowych, rozdzielania zadań między poszczególne węzły i zbierania wyników. Oferuje interfejs dla języka C++. Jest wykorzystywany w większości topowych urządzeń wieloprocessorowych i ta popularność znacząco wpłynęła na jego wybór.

Aplikacja rozproszona - C++11 Wybór padł na ten język ze względu na jego znajomość przez członków grupy oraz interfejs udostępniany przez środowisko MPICH2.

3.2 Narzędzia wykorzystane w projekcie

- Aplikacja webowa PyCharm 5.0.4 - IDE do Pythona, obsługuje Django
- Aplikacja rozproszona - CodeLite 9.1 IDE do C++, wersja na system Linux
- Baza danych - developer do MySQL
- Organizacja pracy - Trello (<https://trello.com/>)
- Hosting plików - GitHub (<https://github.com/>)

4 Aktualny stan rynku

GGNFS (GPL'd implementation of General Number Field Sieve)

Aktywny rozwój. faktoryzacja liczb do 180 znaków, średnio do 140. Kilka większych liczb też. W większości przypadków program GGNFS jest stabilny dla liczb składających się do 150-160 znaków. Posiada bugi. Nie jest czarna skrzynka, trzeba mieć odpowiednią wiedzę, żeby go używać.

Cunningham Project Projekt faktoryzujący liczby $b^n + 1$ dla $b=2,3,5,6,7,10,11,12$ i duże n . [3]

RSA Factoring Challenge Zawody zorganizowane przez RSA Security. Otwarte zawody dla wszystkich mające na celu zwiększyć zainteresowanie faktoryzacją liczb. Opublikowana została lista pseudopierwszych liczb (rozkładających się na dokładnie dwa czynniki), nazwanych liczbami RSA. Za rozłożenie niektórych z nich wyznaczono pieniężną nagrodę. Najmniejsza z nich, 100-cyfrowa liczba RSA-100 została rozłożona w ciągu kilku dni, ale większość do dziś pozostaje niezłamana. Zawody miały na celu śledzenie rozwoju możliwości komputerów w faktoryzacji. Jest to niezwykle istotne przy wyborze długości klucza w szyfrowaniu asymetrycznym metodą RSA. Postęp w łamaniu kolejnych liczb powinien zdradzać jakie długości klucza można jeszcze uznawać za bezpieczne. [4]

5 Projekt techniczny

5.1 Klaster obliczeniowy

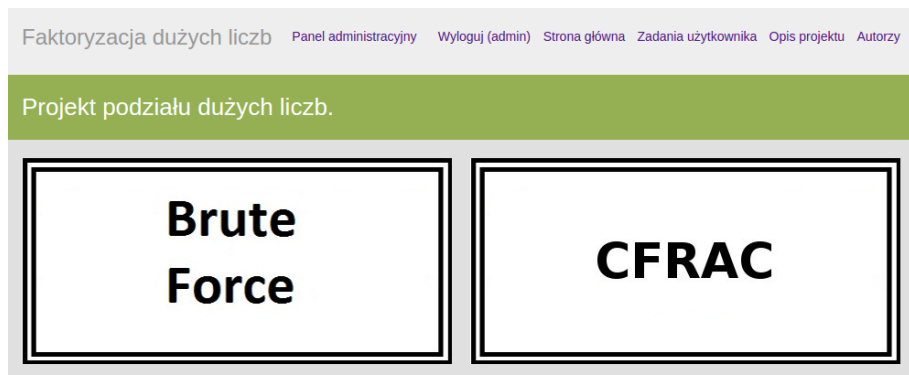
Komunikacja

Algorytm Brute Force

Algorytm CFRAC

5.2 Strona internetowa

Dostępna w sieci Internet strona WWW finalnie umożliwia rejestrację i logowanie użytkowników oraz zlecanie zadań, czyli liczb do faktoryzacji. Do tego użytkownik ma wgląd w historię zleconych przez niego zadań wraz z ich rozwiązaniami. Poza funkcjonalnościami dla zwykłego użytkownika strona umożliwia także działania administracyjne poprzez panel administracyjny. Do zadań tych należą takie rzeczy jak usuwanie zleconych zadań, nadawanie im priorytetu czy też zarządzanie użytkownikami. Strona współpracuje z bazą danych MySQL. Jest ona dostępna pod adresem <http://156.17.235.48/>.



Ostatnie 10 zadań:

Number	User	Algorithm	Date	State	Priority
128	admin	CFRAC	May 24, 2016	2	0
123458756	admin	Brute Force	May 24, 2016	2	0
76348732	admin	Brute Force	May 24, 2016	2	0
126	test	Brute Force	May 24, 2016	2	0
21212121	test	Brute Force	May 23, 2016	2	0
21212121	test	Brute Force	May 23, 2016	2	0
1000000000	test	Brute Force	May 23, 2016	2	0
3727481741	test	CFRAC	May 23, 2016	2	0
3727481741	test	Brute Force	May 23, 2016	2	0
1111	test	Brute Force	May 23, 2016	2	0

Użytkownicy:

Login	Data dołączenia
admin	May 4, 2016, 7:49 p.m.
test	May 4, 2016, 7:50 p.m.

Rysunek 2: Strona główna.

Stronę wykonano przy użyciu technologii Python Django[1] i edytora tekstowego Sublime[5] wraz z dodatkiem Anaconda Python IDE[6]. Do komunikacji z bazą danych wykorzystano standardową bibliotekę do połączenia z MySQL dostępną w Django. Baza danych została stworzona przy pomocy ORM[7] także standardowego dla Django. Przykładowy kod dla encji Task poniżej.

```
class Task(models.Model):
    CANCELLED_STATUS = -1
    UNDONE_STATUS = 0
    WORKING_STATUS = 1
    DONE_STATUS = 2
    STATUS_CHOICES = (
        (CANCELLED_STATUS, "Cancelled"),
        (UNDONE_STATUS, "Undone"),
        (WORKING_STATUS, "Working"),
        (DONE_STATUS, "Done")
    )

    number_to_factor = models.CharField(max_length=200)
    job_date = models.DateField(default=timezone.now)
    state = models.IntegerField(choices=STATUS_CHOICES,
                               default=UNDONE_STATUS)
    priority = models.IntegerField(default=0)
    result = models.CharField(max_length=200)
    user = models.ForeignKey(User, on_delete=models.CASCADE)
```

```
algorithm = models.ForeignKey(Algorithm, on_delete=models.CASCADE)

def __str__(self):
    return str(self.number_to_factor)
```

Kolejne kolumny danej tabeli w bazie danych są tworzone poprzez definiowanie pól klas dziedziczących po `models.Model`[8], tak jak powyżej na przykład kolumna `result` jest polem tekstowym o długości maksymalnej 200 znaków.

W celu utworzenia nowej podstrony definiuje się klasy dziedziczące po klasie `View` z Django. Poniżej znajduje się kod dla widoku podstrony z zadaniami użytkownika.

```
class UserView(LoginRequiredMixin, View):
    template_name = 'FactorerMain/userview.html'

    def get(self, request, *args, **kwargs):
        tasks = Task.objects.filter(user=request.user.id)[::-1]

        context = {'tasks': tasks}

    return render(request, self.template_name, context)
```

W tym przypadku `UserView` dziedziczy również po klasie `LoginRequiredMixin` odpowiadającej za autoryzację, dzięki czemu do takiej podstrony ma dostęp tylko użytkownik zalogowany. W zmiennej `template_name` wskazano na szablon strony opisany w języku HTML. Następnie w metodzie `get` pobierane są wyniki z bazy danych przy pomocy metody `Task.objects.filter(user=request.user.id)[::-1]`. Zwróci to z encji `Task` wyniki należące do aktualnego użytkownika posortowane w od najnowszych do najstarszych. Na końcu zwracany jest szablon oraz kontekst, w tym przypadku zadania użytkownika, poprzez metodę `render`.

Można tak pobrane dane z bazy użyć w HTML i wyświetlić je na stronie. Kod wygląda następująco:

```
<h2>Moje zadania:</h2>
{% if tasks %}
    {% for task in tasks %}
        <details>
            <summary><p>Liczba <b>{{ task.number_to_factor }}</b> przy
                użyciu {{ task.algorithm }} dnia {{ task.job_date }} jest w
                stanie {{ task.state }}</p></summary>
            {% ifequal task.state 2 %}
                {{ task.result }}
            </p>
            {% endifequal %}
        </details>
    </br></br>
```

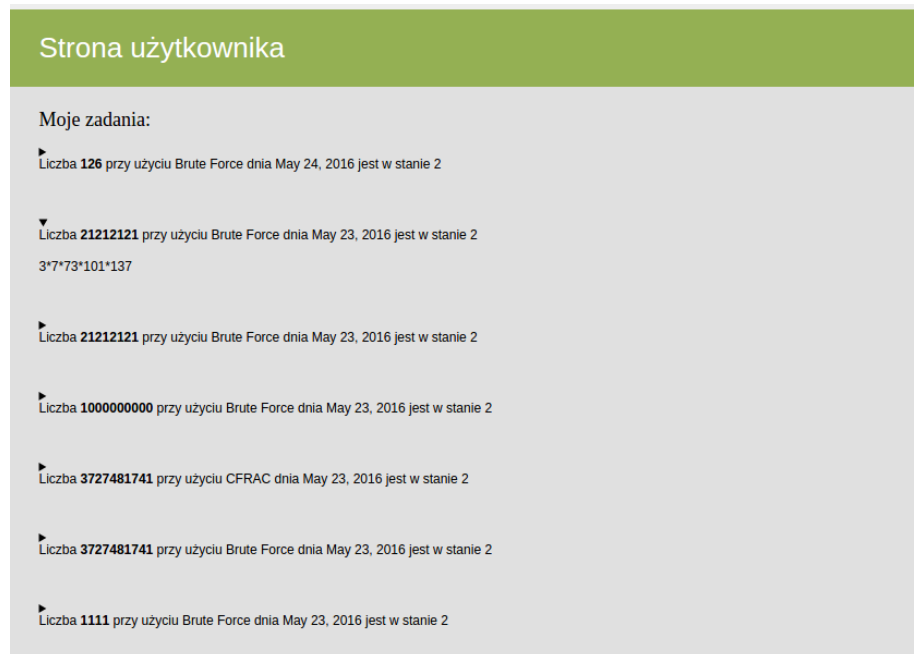


```

    {% endfor %}
{% else %}
    <p>No tasks to show.</p>
{% endif %}

```

Pierw sprawdzane jest czy w kontekście została przekazana zmienna `tasks`. Następnie przy pomocy pętli `for` przechodzi się przez kolejne elementy tabeli i drukuje jej elementy przy pomocy składni `[tabela].[element]`. Przykład uzyskanych z bazy danych wyników na stronie pokazano na rysunku poniżej.



Rysunek 3: Podstrona z zadaniami użytkownika.

W Django zazwyczaj definiuje się jeden bazowy szablon w postaci pliku HTML, a następnie na podstronach zastępuje się tylko wyróżnione bloki strony w sposób wyspecyfikowany dla danej podstrony. Uzyskuje się to przez rozszerzanie szablonów. Dla przykładu w główny pliku HTML znajduje się następujący fragment:

```

<div id="portfolio-samples">
  <div class="wrap clearfix">
    {% block gallery %}
    {% endblock %}
  </div><!-- // end .wrap -->
</div><!-- // end #portfolio-samples -->

```

W tym kodzie zdefiniowano blok gallery, można go zastąpić przy pomocy poniższego fragmentu na innej stronie.

```
{% extends 'base.html' %}

{% block gallery %}
<div class="portfolio-item fl col-2">
  <a href={% url 'bruteforce' %}></a>
    <div class="info">Brute force</div>
</div><!-- // end .portfolio-item -->
<div class="portfolio-item fl push-2 col-2">
  <a href={% url 'cfrac' %}></a>
    <div class="info">CFRAC</div>
</div><!-- // end .portfolio-item -->
{% endblock %}
```

Poza blokiem gallery reszta strony będzie taka jak zdefiniowano w pliku base.html. Strona projektu składa się z jednego bazowego szablonu i rozszerzających go kilku podstron.

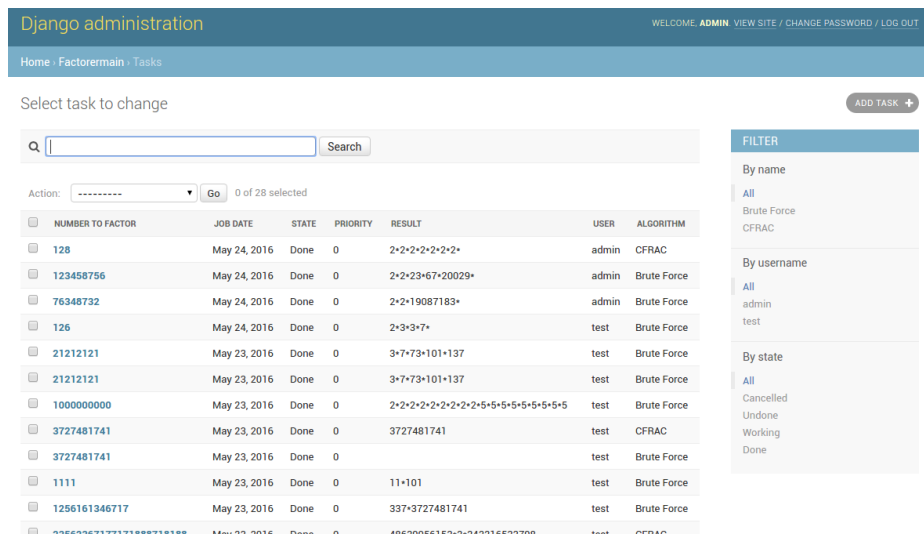
W celu połączenia wszystkiego należy zdefiniować jeszcze w pliku urls.py odwzorowania adresów url dla odpowiednich stron, dzięki czemu Django uruchamia odpowiednie klasy widoków dla zadanego adresu.

```
urlpatterns = [
    url(r'^$', views.IndexView.as_view(), name='index'),
    url(r'^userview/$', views.UserView.as_view(), name='userview'),
    url(r'^about/$', views.AboutView.as_view(), name='about'),
    url(r'^creators/$', views.CreatorsView.as_view(), name='creators'),
    url(r'^bruteforce/$', views.BruteforceView.as_view(),
        name='bruteforce'),
    url(r'^cfrac/$', views.CFRACView.as_view(), name='cfrac'),
    url(r'^login/$', 'django.contrib.auth.views.login'),
    url(r'^logout/$', 'django.contrib.auth.views.logout'),
    url(r'^register/$', views.register, name='register'),
    url(r'^success_register/$', views.SuccessRegisterView.as_view(),
        name='success_register'),
    url(r'^admin/', include(admin.site.urls)),
]
```

Przy tak zdefiniowanych adresach odnosi się do nich w HTML poprzez następującą linię kodu:

```
<a href="{% url 'index' %}">Strona główna</a>
```

Do zadań administracyjnych został wykorzystany standardowy panel Django administration skonfigurowany dla potrzeb projektu.



Rysunek 4: Panel administracyjny Django.

Panel modyfikuje się poprzez edycję pliku admin.py.

```
class TaskAdmin(admin.ModelAdmin):
    list_display = ('number_to_factor', 'job_date', 'state', 'priority',
                   'result', 'user', 'algorithm')
    search_fields = ['user__username']
    list_filter = ('algorithm__name', 'user__username', 'state')

admin.site.register(Task, TaskAdmin)
```

Klasa musi dziedziczyć po klasie admin.ModelAdmin, następnie można zdefiniować sposób wyświetlania danych, pola, po których ma być realizowane wyszukiwanie, czy też listę filtrów. Na końcu trzeba zarejestrować taką klasę dla danej encji.

6 Dokumentacja powykonawcza (instalacyjna)

6.1 Uruchamianie strony

W celu uruchomienia strony potrzebny jest serwer wraz z zainstalowaną powłoką Python 3.4. Do tego należy zainstalować moduł Django w wersji 1.8.7. Oprócz powyższych należy też mieć serwer bazy danych, w przypadku tego projektu jest to serwer MySQL. Po spełnieniu tych wymagań można przejść do uruchamiania strony.

Pierwszym krokiem jest ustawienie w opcjach strony w pliku settings.py

danych dostępowych do bazy danych.

```
DATABASES = {
    'default': {
        'ENGINE': 'django.db.backends.mysql',
        'NAME': 'factorDB',
        'USER': 'projekt',
        'PASSWORD': 'projekt',
        'HOST': '156.17.235.48',
        'PORT': '3306',
    }
}
```

Następnym krokiem jest wypełnienie bazy danych odpowiednimi strukturami stworzonymi wcześniej w Python używając rozwiązania typu ORM. Służy do tego skrypt `manage.py` znajdujący się w folderze z plikami źródłowymi strony. Należy zrobić migrację bazy danych dla naszej aplikacji (nazwa to `FactorerMain`). Osiąga się to przez wywołanie **`python manage.py makemigrations FactorerMain`**. Następnie koniecznym jest wykonać utworzone migrację poprzez **`python manage.py migrate`**. Dzięki tym dwóm komendom baza danych zostanie wypełniona odpowiednio zdefiniowanymi strukturami. Ostatnim krokiem jest uruchomienie serwera, osiąga się to przez użycie kolejnej komendy skryptu `manage.py` - **`python manage.py runserver`**.

```
$ python manage.py runserver
Performing system checks...

System check identified no issues (0 silenced).

June 08, 2016 - 16:24:25
Django version 1.8.7, using settings 'Factorer.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CONTROL-C.
```

Jak widać na powyższym wyciągu z terminala serwer został uruchomiony.

7 Istotne elementy kodu z komentarzem

8 Przykładowe wyniki badań efektywności programu równoległego

9 Wnioski

Literatura

- [1] *Dokumentacja Django*. <https://docs.djangoproject.com/en/1.8/>, 2016
- [2] *Dokumentacja MPICH2*. <http://www.mpich.org/documentation/guides/>, 2016.
- [3] *Cunningham Project*. <http://homes.cerias.purdue.edu/~ssw/cun/>
- [4] *RSA Factoring Challenge*. http://pl.wikipedia.org/wiki/RSA_Factoring_Challenge
- [5] *Sublime Text Editor*. <https://www.sublimetext.com/>
- [6] *Anaconda Python IDE* <http://damnwidget.github.io/anaconda/>
- [7] *ORM - object-relational mapping*. https://pl.wikipedia.org/wiki/Mapowanie_obiektowo-relacyjne
- [8] *Django Models documentation*. <https://docs.djangoproject.com/en/1.9/topics/db/models/>