

# Facebook i R są teraz znajomymi 'Lubię to' czyli jak pobierać dane z facebooka

Łukasz Wawrowski

Uniwersytet Ekonomiczny w Poznaniu

14.03.2014

## Facebook

serwis społecznościowy, w ramach którego zarejestrowani użytkownicy mogą tworzyć sieci i grupy, dzielić się wiadomościami i zdjęciami oraz korzystać z aplikacji, będących własnością Facebook, Inc. z siedzibą w Menlo Park. (Wikipedia 2014)

## Facebook

serwis społecznościowy, w ramach którego zarejestrowani użytkownicy mogą tworzyć sieci i grupy, dzielić się wiadomościami i zdjęciami oraz korzystać z aplikacji, będących własnością Facebook, Inc. z siedzibą w Menlo Park. (Wikipedia 2014)

## Historia

- 4 lutego 2004 — utworzenie projektu TheFacebook
- marzec 2008 — Mark Zuckerberg zostaje najmłodszym miliarderem świata
- maj 2008 — ukazuje się polska wersja językowa serwisu
- wrzesień 2009 — Facebook przestał przynosić straty
- październik 2010 — na ekrany kin wchodzi film „The Social Network”
- 18 maja 2012 — Facebook debiutuje na amerykańskiej giełdzie

# Facebook w liczbach

Liczba użytkowników ogółem:	1 310 000 000
Liczba użytkowników mobilnych:	680 000 000

# Facebook w liczbach

Liczba użytkowników ogółem:	1 310 000 000
Liczba użytkowników mobilnych:	680 000 000
Średni wiek użytkownika:	22 lata
Średni czas spędzany na facebooku:	18 minut
Średnia liczba znajomych:	130

# Facebook w liczbach

Liczba użytkowników ogółem:	1 310 000 000
Liczba użytkowników mobilnych:	680 000 000
Średni wiek użytkownika:	22 lata
Średni czas spędzany na facebooku:	18 minut
Średnia liczba znajomych:	130
Odsetek osób w wieku 18–34 lata, które sprawdzają facebooka po obudzeniu się	48%

# Facebook w liczbach

Liczba użytkowników ogółem:	1 310 000 000
Liczba użytkowników mobilnych:	680 000 000
Średni wiek użytkownika:	22 lata
Średni czas spędzany na facebooku:	18 minut
Średnia liczba znajomych:	130
Odsetek osób w wieku 18–34 lata, które sprawdzają facebooka po obudzeniu się	48%

Rozmiar zgromadzonych danych: 180 petabajtów

# Facebook w liczbach

Liczba użytkowników ogółem:	1 310 000 000
Liczba użytkowników mobilnych:	680 000 000
Średni wiek użytkownika:	22 lata
Średni czas spędzany na facebooku:	18 minut
Średnia liczba znajomych:	130
Odsetek osób w wieku 18–34 lata, które sprawdzają facebooka po obudzeniu się	48%

Rozmiar zgromadzonych danych: 180 petabajtów

Co 24 godziny przybywa ponad 0,5 petabajta

180 PB = 184 320 TB = 188 743 680 GB

180 PB = 184 320 TB = 188 743 680 GB

- 27 648 000 000 plików mp3 o rozmiarze 7 MB

180 PB = 184 320 TB = 188 743 680 GB

- 27 648 000 000 plików mp3 o rozmiarze 7 MB
- 22 118 400 godzin (2 524 lat) filmów

$$180 \text{ PB} = 184\,320 \text{ TB} = 188\,743\,680 \text{ GB}$$

- 27 648 000 000 plików mp3 o rozmiarze 7 MB
- 22 118 400 godzin (2 524 lat) filmów
- wieża ułożona z dysków twardych o pojemności 1 TB i wysokości 4 cm miałaby 7372,8 metrów wysokości



$$180 \text{ PB} = 184\,320 \text{ TB} = 188\,743\,680 \text{ GB}$$

- 27 648 000 000 plików mp3 o rozmiarze 7 MB
- 22 118 400 godzin (2 524 lat) filmów
- wieża ułożona z dysków twardych o pojemności 1 TB i wysokości 4 cm miałaby 7372,8 metrów wysokości



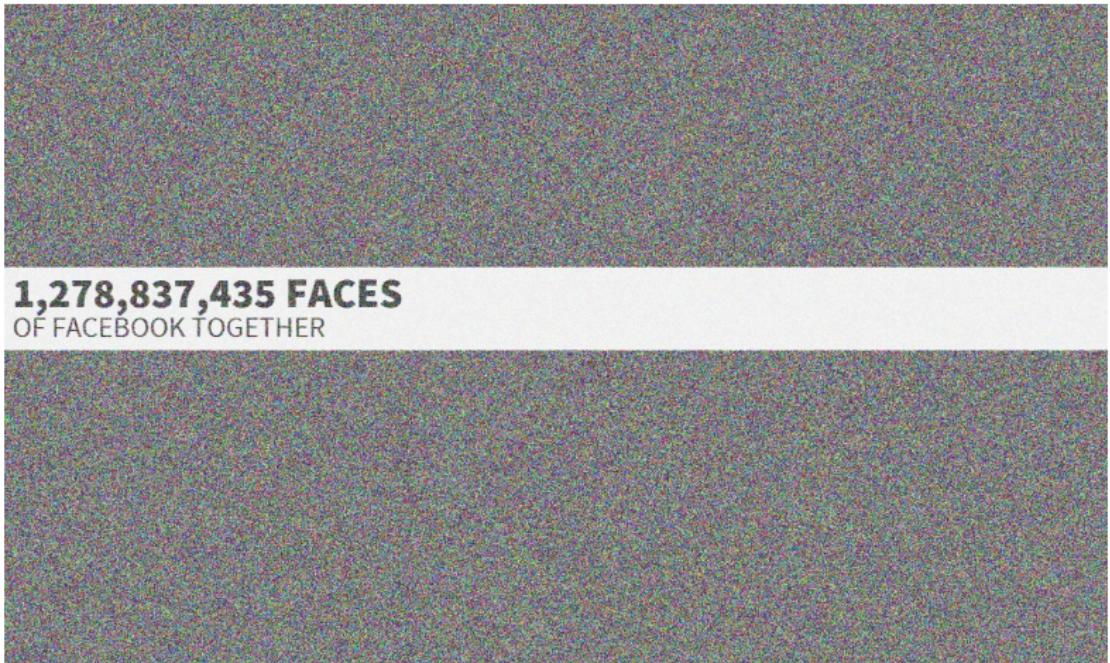
czyli prawie tyle co Sherpi Kangri — 74 szczyt Ziemi

# Paul Butler — „Visualizing Friendships”



<https://www.facebook.com/notes/facebook-engineering/visualizing-friendships/469716398919>

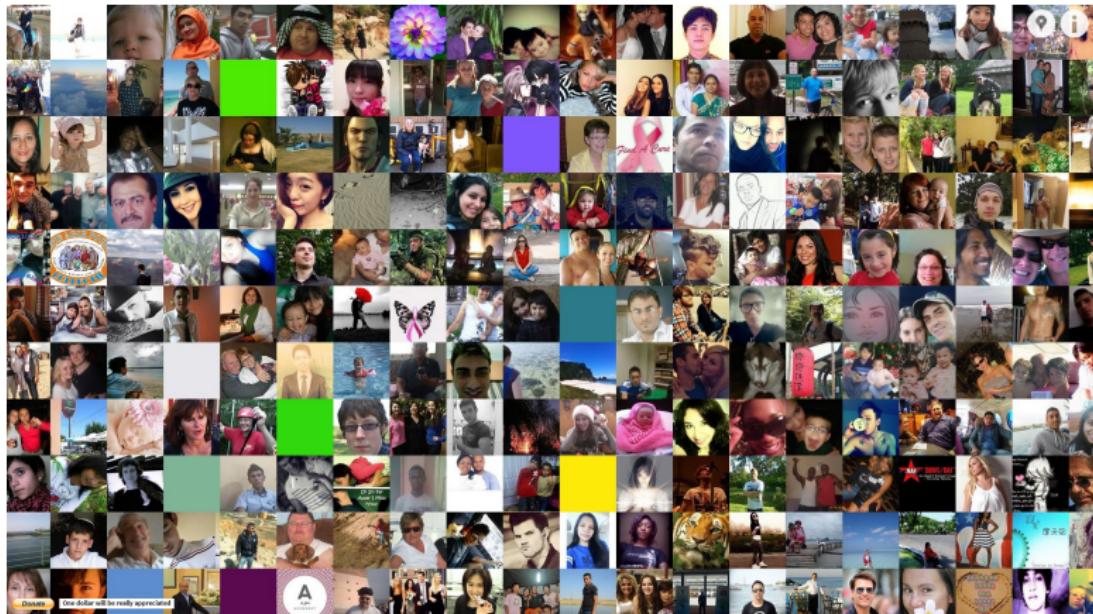
# Natalia Rojas — „The Faces of Facebook”



**1,278,837,435 FACES**  
OF FACEBOOK TOGETHER

<http://thefacesoffacebook.com/>

# Natalia Rojas — „The Faces of Facebook”

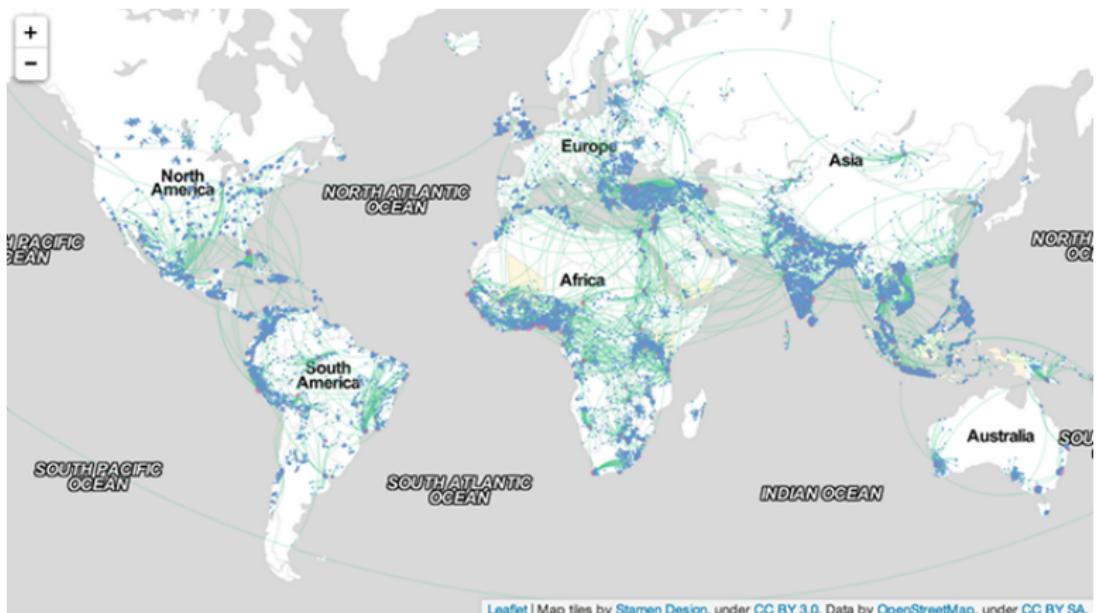


<http://thefaceoffacebook.com/>

Donate One dollar will be really appreciated



# Aude Hofleitner — „Coordinated Migration“



<https://www.facebook.com/notes/facebook-data-science/coordinated-migration/10151930946453859>

[Twórcy aplikacji](#) [Aplikacje](#) [Produkty](#) [Dokumenty](#) [Narzędzia](#) [Wsparcie](#) [Szukaj w dokumentach](#)

### Graph Explorer



Explore the Graph API from your browser. Test permissions, query data and learn more about the Graph API syntax.

### Panel statystyk



Insights provides you with analytics on your Facebook Pages, Facebook Apps and Social Plugins on your website.

### Przeglądarka obiektów



The Object Browser lets you browse and edit Open Graph objects that you've created using the Object API, and create new objects.

### Menedżer reklam



Use the Ads Manager to create ads, setup sponsored stories, manage your ads budget and measure the impact of your campaigns.

### Narzędzie do debugowania protokołu Open Graph



Helps you debug your Open Graph pages. Start here if you're having problems with the Like button or our other Social Plugins.

### Tokeny dostępu



Get access tokens for your apps for easy testing. Use the Inter to debug access tokens, test token expiration and check user permissions.

### Konsola testowa Javascript



A simple way to learn, test and debug the Facebook SDK for Javascript. It also includes a large number of working examples.

#### Facebook Developers

<b>Produkty</b> Logowanie za pośrednictwem Facebooka Udostępnianie Analizuj składnię Gry Reklamy aplikacji	<b>Pakiet SDK</b> iOS SDK Android SDK JavaScript SDK PHP SDK Unity SDK	<b>Narzędzia</b> Graph API Explorer Open Graph Debugger Przeglądarka obiektów Konsola testowa Javascript Statystyki na Facebooku	<b>Pomoc</b> Status platformy Grupa twórców Preferowani twórcy aplikacji Błędy	<b>Nowości</b> Blog Mapa strategii dla twórców Prezentacje
---	---	---	--	---

[Facebook © 2014 - Polski](#) [Informacje](#) [Utwórz reklamę](#) [Praca](#) [Polityka platformy](#) [Polityka prywatności](#)

<https://developers.facebook.com/tools/>

# Graph API Explorer

Twórcy aplikacji Aplikacje Produkty Dokumenty Narzędzia Wsparcie Szukaj w dokumentach

## Graph API Explorer

Aplikacja: [?] Graph API Explorer Język: [?] English (US) ▾

Token dostępowy: CAACEdEose0cBAPHcjkHauwbynFn3QDTul41PAChzdgAmF X Debuguj Otrzymaj token dostępowy Uzyskaj token aplikacji

Graph API FQL Query

GET → /1836011134?fields=id,name,gender,likes.fields(name)

Wyslij

Dowiedz się więcej o składni interfejsu Graph API

Node: 1836011134

id  
 name  
 gender  
 likes →  
 name  
Search for a field

Search for a field

```
[{"id": "1836011134", "name": "Łukasz Nawrocki", "gender": "male", "likes": {"data": [{"id": "8394258414", "name": "Facebook Data Science"}, {"id": "220046204721532", "name": "Nie czajesz? Nie idę z Tobą do kózka"}, {"id": "324452747410411", "name": "Erik Witsoe Photography"}, {"id": "500926169152154", "name": "Na Strąży Sondaży"}, {"id": "460753360614176", "name": "Gnieźnieński fyrtel"}, {"id": "60667806884", "name": "digart.pl"}, {"id": "689634084412587", "name": "Praca jest"}, {"id": "888888888888888", "name": "EKN Portfolio - Studenckie Koło Naukowe UEB"}]
```

Odpowiedź nadieszła w ciągu 118 ms Save Session

<https://developers.facebook.com/tools/explorer/>

## Token dostępowy

- CAACEdEose0cBAIW DvXwN6ZC73ABSEF6x5nFjAaI5OibvY  
ZCLJQCXcCrZAbSdt9K5WSdQEKI8RdypNR3BZA9yoc9KX  
tZBK aUVZA6JrSqR7sCfyDxk1sffwxv1ZBbriytRFfaPyqP4AZD  
RVNZAp glnouzzs1tdKsLuflnxM1FdUBqsA0qrROn0...

## Token dostępowy

- CAACEdEose0cBAIW DvXwN6ZC73ABSEF6x5nFjAaI5OibvY ZCLJQCXcCrZAbSdt9K5WSdQEKI8RdypNR3BZA9yocy9KX tZBKaUVZA6JrSqR7sCfyDxk1sffwxv1ZBbriytRFfaPyqP4AZD RVNZAp glnouzzs1tdKsLuflnxM1FdUBqsA0qrROno0...
- Zawiera informacje o właścielie

# Token dostępowy

- CAACEdEose0cBAIW DvXwN6ZC73ABSEF6x5nFjAaI5OibvY ZCLJQCXcCrZAbSdt9K5WSdQEKI8RdypNR3BZA9yoc9KX tZBKaUVZA6JrSqR7sCfyDxk1sffwxv1ZBbriytRFfaPyqP4AZD RVNZAp gl nouzzs1tdKsLuflnxM1FdUBqsA0qrROno0...
- Zawiera informacje o właściwielu
- oraz o danych, które można pobierać. Wyróżnia się trzy obszary pozwoleń:
  - informacje o użytkowniku (np. edukacja, zainteresowania),
  - informacje o znajomych (np. urodziny, check-iny),
  - rozszerzone (np. publikowanie statusu).

# Token dostępowy

- CAACEdEose0cBAIW DvXwN6ZC73ABSEF6x5nFjAaI5OibvY ZCLJQCXcCrZAbSdt9K5WSdQEKI8RdypNR3BZA9yoc9KX tZBKaUVZA6JrSqR7sCfyDxk1sffwxv1ZBbriytRFfaPyqP4AZD RVNZAp gl nouzzs1tdKsLuflnxM1FdUBqsA0qrROno0...
- Zawiera informacje o właściwielu
- oraz o danych, które można pobierać. Wyróżnia się trzy obszary pozwoleń:
  - informacje o użytkowniku (np. edukacja, zainteresowania),
  - informacje o znajomych (np. urodziny, check-iny),
  - rozszerzone (np. publikowanie statusu).
- Przechowuje także datę wygaśnięcia:
  - token ważny jest przez godzinę.

# XML vs JSON

XML, Extensible Markup Language — uniwersalny język formalny przeznaczony do reprezentowania różnych danych w strukturalizowany sposób.

JSON, JavaScript Object Notation — lekki format wymiany danych komputerowych.

# XML vs JSON

XML, Extensible Markup Language — uniwersalny język formalny przeznaczony do reprezentowania różnych danych w strukturalizowany sposób.

JSON, JavaScript Object Notation — lekki format wymiany danych komputerowych.

```
<?xml version="1.0" encoding="UTF-8" ?>
<id>1836011134</id>
<name>Łukasz Wawrowski</name>
<work>
<employer>
<id>435082969934599</id>
<name>Urząd Statystyczny</name>
</employer>
<location>
<id>113787108631661</id>
<name>Poznań</name>
</location>
<position>
<id>108465685875101</id>
<name>Specjalista</name>
</position>
<start_date>2011-09-01</start_date>
</work>
```

```
{
    "id": "1836011134",
    "name": "Łukasz Wawrowski",
    "work": [
        {
            "employer": {
                "id": "435082969934599",
                "name": "Urząd Statystyczny"
            },
            "location": {
                "id": "113787108631661",
                "name": "Poznań"
            },
            "position": {
                "id": "108465685875101",
                "name": "Specjalista"
            },
            "start_date": "2011-09-01"
        }
    ]
}
```

## Pakiet rjson

`fromJSON` — konwertuje obiekt JSON do obiektu R

`newJSONParser` — konwertuje zbiór obiektów JSON do obiektu R

`toJSON` — konwertuje obiekt R do obiektu JSON

## Pakiet rjson

`fromJSON` — konwertuje obiekt JSON do obiektu R  
`newJSONParser` — konwertuje zbiór obiektów JSON do obiektu R  
`toJSON` — konwertuje obiekt R do obiektu JSON

```
token="CAACEdEose0cBAMKJg8vUBvHZC5PEHmTGCFRbzuVDfiq
VK5T3P5m4N4w5H6pWujCNVFSZAbYCdjmMpJCpNgTLDMZBmNydgE"
```

```
json_file=paste("https://graph.facebook.com/me/
?fields=id,name&access_token=", token, sep="")
```

```
json_data=fromJSON(paste(readLines(json_file),
collapse=""))
json_data
```

```
$id
[1] "1836011134"
```

```
$name
[1] "Łukasz Wawrowski"
```

## Pakiet rjson

```
token="CAACEdEose0cBAMKJg8vUBvHZC5PEHmTGCFRbzuVDfiq  
VK5T3P5m4N4w5H6pWujCNVFSZAbYCdjmMpJCpNgTLDMZBmNydgE"  
  
json_file=paste("https://graph.facebook.com/me/  
friends?fields=id,location&access_token=", token, sep="")  
  
json_data=fromJSON(paste(readLines(json_file),  
collapse=""))  
json_data  
  
$data[[310]]  
$data[[310]]$id  
[1] "100001641304264"  
  
$data[[310]]$location  
$data[[310]]$location$id  
[1] "106029346094605"  
  
$data[[310]]$location$name  
[1] "Kraków, Poland"
```

# Pakiet Rfacebook

`getCheckins` — pobiera informacje o 'zameldowaniach' znajomych

`getFriends` — pobiera informacje o znajomych

`getLikes` — pobiera informacje o polubieniach znajomych

`getNetwork` — zwraca listę wspólnych znajomych lub macierz znajomości

`getNewsfeed` — pobiera aktualności z tablicy

`getPage` — pobiera informacje z publicznych fanpage

`getPost` — pobiera informacje o publicznym poście z listą komentarzy i polubień

`getUsers` — pobiera informacje o użytkowniku

`searchFacebook` — pobiera publiczne statusy zawierające określoną treść

`updateStatus` — ustawia status na profilu

Jak to działa?



# Ocena jakości danych pobranych z facebooka

Cel: stworzyć mapę znajomych według miejsca zamieszkania i pochodzenia.

# Ocena jakości danych pobranych z facebooka

Cel: stworzyć mapę znajomych według miejsca zamieszkania i pochodzenia.

Procent znajomych na facebooku: 100%

# Ocena jakości danych pobranych z facebooka

Cel: stworzyć mapę znajomych według miejsca zamieszkania i pochodzenia.

Procent znajomych na facebooku: 100%

Procent znajomych pobrany przez pakiet: 97%

# Ocena jakości danych pobranych z facebooka

Cel: stworzyć mapę znajomych według miejsca zamieszkania i pochodzenia.

Procent znajomych na facebooku:	100%
Procent znajomych pobrany przez pakiet:	97%
Procent niepustych miejsc zamieszkania:	66%
Procent niepustych miejsc pochodzenia:	62%

# Ocena jakości danych pobranych z facebooka

Cel: stworzyć mapę znajomych według miejsca zamieszkania i pochodzenia.

Procent znajomych na facebooku:	100%
Procent znajomych pobrany przez pakiet:	97%
Procent niepustych miejsc zamieszkania:	66%
Procent niepustych miejsc pochodzenia:	62%
Procent zgeokodowanych miejsc zamieszkania:	66%
Procent zgeokodowanych miejsc pochodzenia:	61%

Google Maps miał problem z *Brójce, Gorzów, Poland*

# Znajomi według miejsca zamieszkania

77 unikalnych miejsc



# Znajomi według miejsca pochodzenia

131 unikalnych miejsc

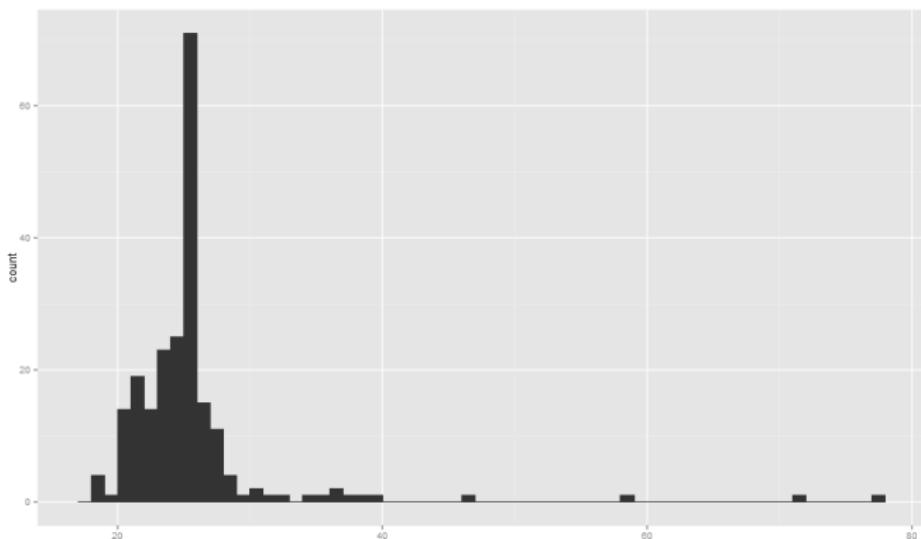


## Znajomi według imion

159 unikalnych imion

# Znajomi według wieku

25 unikalnych wartości



Opracowanie własne.

# Migracje znajomych

55% znajomych miało uzupełnione zarówno miejsce pochodzenia jak i zamieszkania



# Podsumowanie

- wiodącym formatem danych w Internecie staje się JSON
- R bez problemu radzi sobie z tym formatem
- zwykły użytkownik ma dostęp tylko do danych publicznych oraz tych powiązanych z nim
- ustawienia prywatności uniemożliwiają pobranie wszystkich informacji
- braki danych znacznie ograniczają analizowany zbiór danych
- użytkownik sam wpisuje informacje, więc nie ma obowiązku podawać prawdziwych danych
- pomimo braków i błędów dane z facebooka mają olbrzymi potencjał

„Żyjemy w czasach bezprecedensowych możliwości gromadzenia i analizy danych dotyczących szczegółów życia człowieka takich jak:

- wzorce zakupowe,
- styl życia,
- związki i więzi społeczne,
- potrzeby, opinie oraz
- ruchy.”

(Pratesi i in. 2013)

<https://www.facebook.com/sknEstymator>

<https://www.facebook.com/sknEstymator>

Dziękuje za uwagę.