

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE

Fakulta informatiky a informačných technológií

Ilkovičova 2, 842 16 Bratislava 4

Analytický report

Vouky – osobný slovník

Katarína Rafčíková

Študijný program: Informatika 4-ročný

Ročník: 3, Krúžok: Streda 10:00

Predmet: Databázové systémy

Vedúci projektu: Ing. Tomáš Kramár, PhD.

Ak. rok: 2016/2017

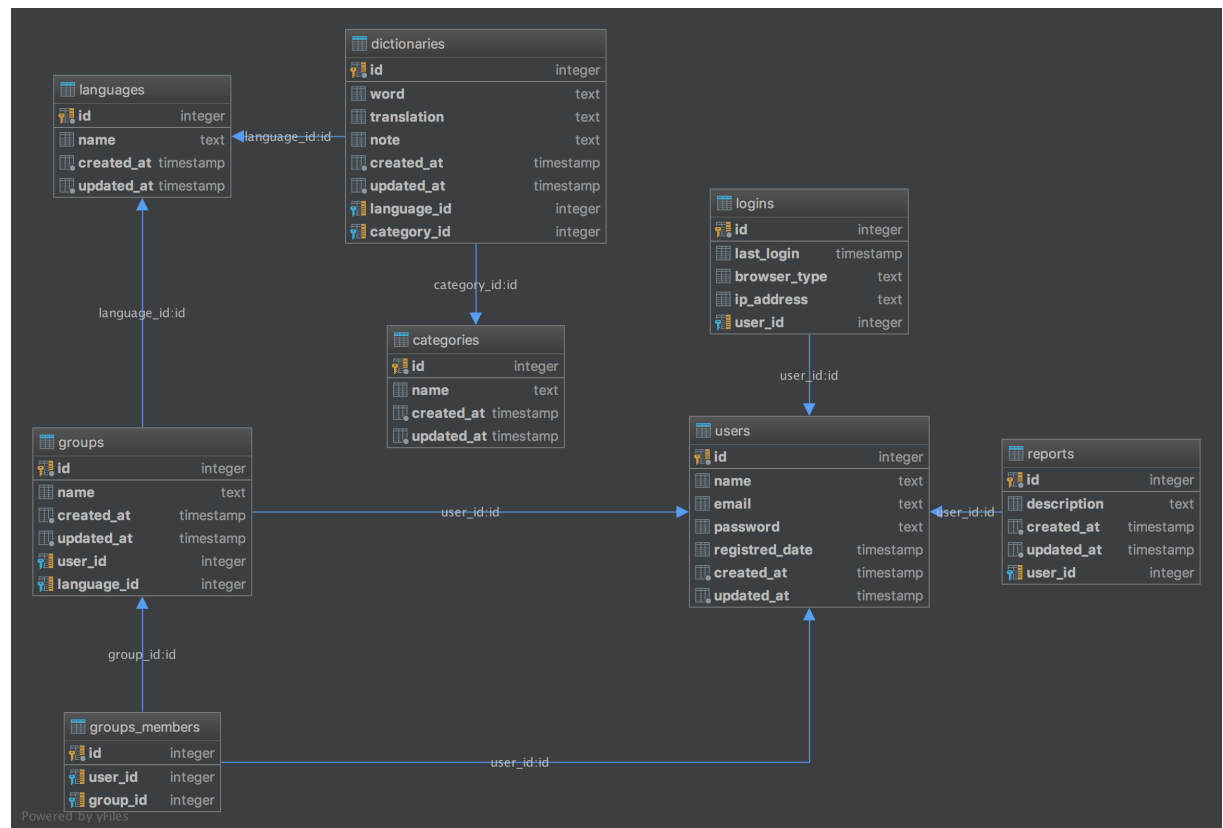
Obsah

1.	Základné informácie o projekte	3
2.	Model.....	3
3.	Generovanie dát.....	3
4.	Prehľad tabuliek.....	4
3.1.	users.....	4
3.1.1	SQL queries.....	4
3.2.	groups	4
3.3.	groups_members	5
3.4.	reports	5
3.5.	logins.....	5
3.5.1	SQL queries.....	5
3.6.	dictionaries	6
3.6.1	SQL queries.....	6
3.7.	languages	6
3.8.	categories.....	7
3.8.1	SQL queries.....	7
5.	Bibliografia	7

1. Základné informácie o projekte

Webová aplikácia, osobný slovník Vouky, je vyvíjaná vo frameworku Rails. Jej úlohou je podporiť výučbu cudzích jazykov a pomôcť študentom s evidenciou slovíčok.

2. Model



Obrázok 1: Fyzický model

3. Generovanie dát

- dáta do tabuliek boli vygenerované náhodne pomocou nástroja Faker [1]
- číselník languages je naplnený dátami z Google Translate [2]
- tabuľka categories je naplnená dátami z ManyThings [3]
- stĺpec browser_type v tabuľke logins je naplnený dátami z Wikipedia [4]

4. Prehľad tabuliek

3.1. users

Počet riadkov	113280
Počet mien (atribút name) vyskytujúcich sa 1x	109396
Počet mien (atribút name) vyskytujúcich sa 2x	1892
Počet mien (atribút name) vyskytujúcich sa 3x	32
Počet mien (atribút name) vyskytujúcich sa 4x	1
Počet emailov (atribút email) vyskytujúcich sa 1x	113136
Počet emailov (atribút email) vyskytujúcich sa 2x	72
Počet hesiel (atribút password) vyskytujúcich sa 1x	113280

3.1.1 SQL queries

```
CREATE TABLE users
(
  id SERIAL NOT NULL
    CONSTRAINT users_pkey
    PRIMARY KEY,
  name text,
  email text,
  password text,
  registred_date timestamp,
  created_at timestamp NOT NULL,
  updated_at timestamp NOT NULL
);
```

3.2. groups

	Min	Max	Avg	Dolný kvartil	Medián	Horný kvartil
user_id	45004	158281	101803.83876	73462.75	101823.5	130441.5
language_id	1	103	52.20535	27	52	78

Počet riadkov	100000
Počet mien (atribút name) vyskytujúcich sa 1x	96939
Počet mien (atribút name) vyskytujúcich sa 2x	1478
Počet mien (atribút name) vyskytujúcich sa 3x	35

3.3. groups_members

	Min	Max	Avg	Dolný kvartil	Medián	Horný kvartil
user_id	45006	158280	101786.21831	73501	101807	130138.25
group_id	168400	268398	218492.08749	193491.75	218660	243451

Počet riadkov	100000
---------------	--------

3.4. reports

	Min	Max	Avg	Dolný kvartil	Medián	Horný kvartil
user_id	45003	158276	101646.06402	73359	101679	129804.25

Počet riadkov	100000
Počet popisov (atribút description) vyskytujúcich sa 1x	99998
Počet popisov (atribút description) vyskytujúcich sa 2x	1

3.5. logins

	Min	Max	Avg	Dolný kvartil	Medián	Horný kvartil
user_id	45003	158282	101816.6759	73610.75	101867	130174

Počet riadkov	100000
Počet riadkov kde browser_type = Internet Explorer	16984
Počet riadkov kde browser_type = Opera	16487
Počet riadkov kde browser_type = Safari	16641
Počet riadkov kde browser_type = Mozilla Firefox	16552
Počet riadkov kde browser_type = Google Chrome	16713
Počet riadkov kde browser_type = Microsoft Edge	16623

3.5.1 SQL queries

```
SELECT count(*) FROM logins WHERE browser_type='Internet Explorer';
```

3.6. dictionaries

Počet riadkov: 1000000

	Min	Max	Avg	Dolný kvartil	Medián	Horný kvartil
user_id	45003	158282	101602.378361	73265	101592	129961
category_id	1	199	100.025684	50	100	150
language_id	1	103	51.980691	26	52	78

Počet riadkov	100000
Počet riadkov kde word= et	44393
Počet riadkov kde word= qui	24269
Počet riadkov kde word= ut	24245
Počet riadkov kde word= aut	23952
Počet riadkov kde word= est	20182
Počet poznámok (atribút note) vyskytujúcich sa 1x	999824
Počet poznámok (atribút note) vyskytujúcich sa 2x	88

3.6.1 SQL queries

```
-- min, max, avg
SELECT min(language_id), max(language_id), avg(language_id)
FROM dictionaries;
```

```
-- dolny kvartil
SELECT percentile_cont(0.25) WITHIN GROUP (ORDER BY
language_id) FROM dictionaries;
```

```
-- median
SELECT percentile_cont(0.5) WITHIN GROUP (ORDER BY
language_id) FROM dictionaries;
```

```
-- horny kvartil
SELECT percentile_cont(0.75) WITHIN GROUP (ORDER BY
language_id) FROM dictionaries;
```

3.7. languages

Počet riadkov	103
Všetky hodnoty atribútu name sú unikátne - číselník	

3.8. categories

Počet riadkov	199
Počet názvov (atribút name) vyskytujúcich sa 1x	199

3.8.1 SQL queries

```
SELECT count(*)  
FROM  
    (SELECT u.name, count(u.name)  
    FROM categories u  
    GROUP BY u.name  
    HAVING count(u.name) = 1  
    ORDER BY count(u.name) DESC)  
    AS sub
```

5. Bibliografia

- [1] Faker. 2017. [Online] <http://faker.readthedocs.io/en/master/>
- [2] Google Translate. 2017. [Online] <https://translate.google.com/>
- [3] ManyThings, Comprehensive English Vocabulary Word Lists. 2017. [Online] <http://www.manythings.org/vocabulary/lists/c/>
- [4] Wikipedia. List of web browsers. 2017. [Online] https://en.wikipedia.org/wiki/List_of_web_browsers