

ČVUT, FAKULTA INFORMAČNÍCH TECHNOLOGIÍ  
VYHLEDÁVÁNÍ NA WEBU A V MULTIMEDIÁLNÍCH DATABÁZÍCH  
LETNÍ SEMESTR 2019/2020  
ZÁVĚREČNÁ ZPRÁVA K PROJEKTU

---

# LSI vektorový model

---

David Mašek a Kristýna Klesnilová

4. května 2020

## OBSAH

<b>1</b>	<b>Popis projektu</b>	<b>3</b>
<b>2</b>	<b>Způsob řešení</b>	<b>3</b>
2.1	Preprocessing . . . . .	3
2.2	Implementace LSI . . . . .	3
2.3	Vyhodnocení dotazu . . . . .	3
<b>3</b>	<b>Implementace</b>	<b>3</b>
3.1	Identify the author of Equation 3.1 below and briefly describe it in English. . . . .	3
3.2	Try to make sense of some more equations. . . . .	3
<b>4</b>	<b>Příklad výstupu</b>	<b>4</b>
4.1	Bullet Point List . . . . .	4
4.2	Numbered List . . . . .	4
<b>5</b>	<b>Experimentální sekce</b>	<b>4</b>
5.1	The table above shows the nutritional consistencies of two sausage types. Explain their relative differences given what you know about daily adult nutritional recommendations. . . . .	4
<b>6</b>	<b>Diskuze</b>	<b>4</b>
6.1	How many luftballons will be output by the Listing 1 above? . . . . .	5
6.2	Identify the regular expression in Listing 1 and explain how it relates to the anti-war sentiments found in the rest of the script. . . . .	5
<b>7</b>	<b>Závěr</b>	<b>5</b>

# 1 POPIS PROJEKTU

V tomto projektu implementujeme LSI vektorový model sloužící k podobnostnímu vyhledávání v databázi textových dokumentů. Tuto funkcionalitu následně vizualizujeme pomocí webového interface, který uživateli umožňuje procházet databázi článků na základě doporučení nejpodobnějších článků k právě čtenému.

V experimentální části projektu jsme se dále zaměřili na:

- Porovnání průchodu pomocí LSI vektorového modelu se sekvenčním průchodem databáze
- Porovnání vlivu LSI na kvalitu výsledků vyhledávání s ohledem na výskyt synonym a homonym
- Vliv různých vnitřních parametrů na výkon algoritmu (změna počtu konceptů, změna počtu extrahovaných termů, použití lemmatizace namísto stemmingu...)

Celý náš projekt je volně dostupný na: (Odkaz na gitlab?).

## 2 ZPŮSOB ŘEŠENÍ

### 2.1 Preprocessing

### 2.2 Implementace LSI

### 2.3 Vyhodnocení dotazu

## 3 IMPLEMENTACE

### 3.1 Identify the author of Equation 3.1 below and briefly describe it in English.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3.1)$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent porttitor arcu luctus, imperdiet urna iaculis, mattis eros. Pellentesque iaculis odio vel nisl ullamcorper, nec faucibus ipsum molestie. Sed dictum nisl non aliquet porttitor. Etiam vulputate arcu dignissim, finibus sem et, viverra nisl. Aenean luctus congue massa, ut laoreet metus ornare in. Nunc fermentum nisi imperdiet lectus tincidunt vestibulum at ac elit. Nulla mattis nisl eu malesuada suscipit.

### 3.2 Try to make sense of some more equations.

$$\begin{aligned} (x+y)^3 &= (x+y)^2(x+y) \\ &= (x^2+2xy+y^2)(x+y) \\ &= (x^3+2x^2y+xy^2) + (x^2y+2xy^2+y^3) \\ &= x^3+3x^2y+3xy^2+y^3 \end{aligned} \quad (3.2)$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

$$A = \begin{bmatrix} A_{11} & A_{21} \\ A_{21} & A_{22} \end{bmatrix} \quad (3.3)$$

Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem.

## 4 PŘÍKLAD VÝSTUPU

### 4.1 Bullet Point List

- First item in a list
  - First item in a list
    - \* First item in a list
    - \* Second item in a list
  - Second item in a list
- Second item in a list

### 4.2 Numbered List

1. First item in a list
2. Second item in a list
3. Third item in a list

## 5 EXPERIMENTÁLNÍ SEKCE

<i>Per 50g</i>	<b>Pork</b>	<b>Soy</b>
Energy	760kJ	538kJ
Protein	7.0g	9.3g
Carbohydrate	0.0g	4.9g
Fat	16.8g	9.1g
Sodium	0.4g	0.4g
Fibre	0.0g	1.4g

Tabulka 5.1: Sausage nutrition.

### 5.1 The table above shows the nutritional consistencies of two sausage types. Explain their relative differences given what you know about daily adult nutritional recommendations.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent porttitor arcu luctus, imperdiet urna iaculis, mattis eros. Pellentesque iaculis odio vel nisl ullamcorper, nec faucibus ipsum molestie. Sed dictum nisl non aliquet porttitor. Etiam vulputate arcu dignissim, finibus sem et, viverra nisl. Aenean luctus congue massa, ut laoreet metus ornare in. Nunc fermentum nisi imperdiet lectus tincidunt vestibulum at ac elit. Nulla mattis nisl eu malesuada suscipit.

## 6 DISKUZE

Listing 1: Luftballons Perl Script.

```
1 #!/usr/bin/perl
2
3 use strict;
4 use warnings;
```

```

5
6 for (1..99) { print $_." Luftballons\n"; }
7
8 # This is a commented line
9
10 my $string = "Hello World!";
11
12 print $string."\n\n";
13
14 $string =~ s/Hello/Goodbye Cruel/;
15
16 print $string."\n\n";
17
18 finale ();
19
20 exit;
21
22 sub finale { print "Fin.\n"; }

```

### 6.1 How many luftballons will be output by the Listing 1 above?

Aliquam arcu turpis, ultrices sed luctus ac, vehicula id metus. Morbi eu feugiat velit, et tempus augue. Proin ac mattis tortor. Donec tincidunt, ante rhoncus luctus semper, arcu lorem lobortis justo, nec convallis ante quam quis lectus. Aenean tincidunt sodales massa, et hendrerit tellus mattis ac. Sed non pretium nibh. Donec cursus maximus luctus. Vivamus lobortis eros et massa porta porttitor.

### 6.2 Identify the regular expression in Listing 1 and explain how it relates to the anti-war sentiments found in the rest of the script.

Fusce varius orci ac magna dapibus porttitor. In tempor leo a neque bibendum sollicitudin. Nulla pretium fermentum nisi, eget sodales magna facilisis eu. Praesent aliquet nulla ut bibendum lacinia. Donec vel mauris vulputate, commodo ligula ut, egestas orci. Suspendisse commodo odio sed hendrerit lobortis. Donec finibus eros erat, vel ornare enim mattis et.

## 7 ZÁVĚR