

**ПРЕЗЕНТАЦІЯ  
МАГІСТЕРСЬКОЇ КВАЛІФІКАЦІЙНОЇ  
РОБОТИ НА ТЕМУ:**

**«ВДОСКОНАЛЕННЯ МЕТОДУ ІДЕНТИФІКАЦІЇ СУБ`ЄКТА З  
ВІДДАЛЕНИМ ДОСТУПОМ НА ОСНОВІ АНАЛІЗУ ЧАСТОТНИХ  
ХАРАКТЕРИСТИК ЙОГО ЛЕКСИКИ»**

**Студента групи УБ-16м  
Сухоплеско Богдана Олександровича**

**к.ф. - м.н., проф., Шиян А.А.**

## **Метою роботи є**

удосконалення методу ідентифікації особи що працює віддалено за частотними характеристиками його лексики.

Для виконання мети роботи поставлено наступні задачі:

- проаналізувати методи та засоби ідентифікації особливості;
- удосконалити метод автентифікації за частотними показниками лексики, що поєднує закон Ціпфа та алгоритм TF-IDF;
- розробити програмний засіб для реалізації методу аналізу частотних характеристик лексики.

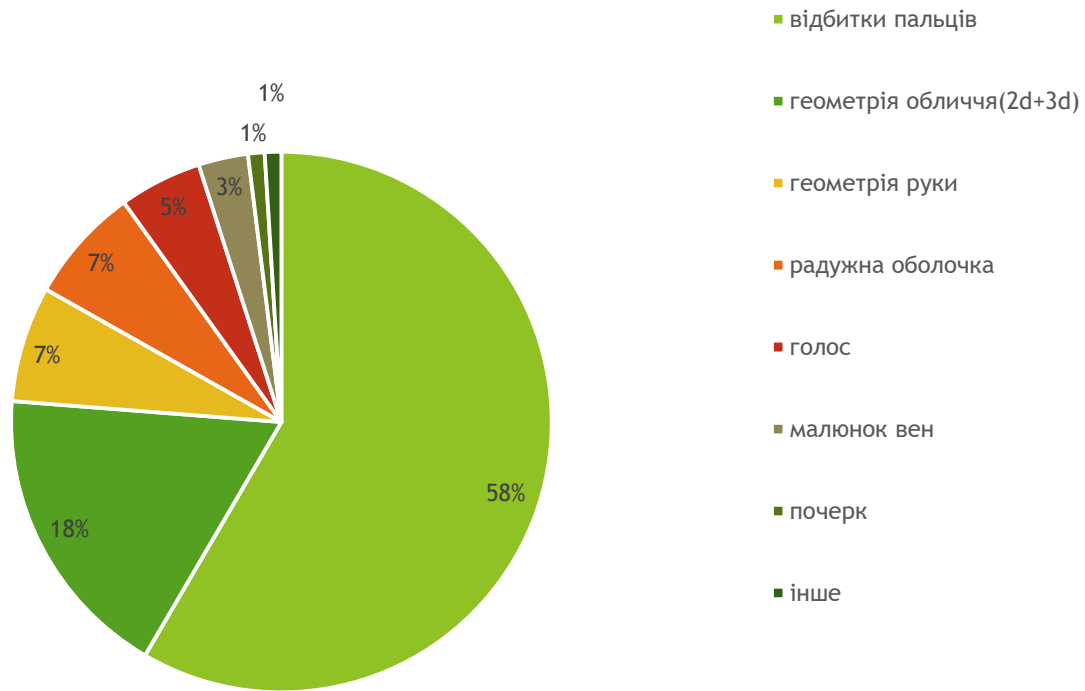
**Об'єктом** дослідження є лексика та лексичний словник людини, які їй притаманні під час щоденної діяльності та в умовах праці.

# Актуальність

З кожним днем сучасні підприємства розвиваються все більше, переважна більшість вже розширюється, створюючи власні філіали в різних містах, країнах. В деяких випадках підприємство настільки виростає, що певні філіали знаходяться навіть на інших континентах. В таких великих підприємствах дуже велика кількість кадрів. Також сучасність дозволяє, і багато роботодавців цим користується, надавати можливість працювати віддалено. Це дає змогу економити на робочих місцях, забезпечити більш комфортні умови для роботи працівників і позбавляє необхідності особистої присутності у випадках коли вона не є обов'язковою.

Також такий ріст підприємств впливає на те, що деякі підрозділи можуть знаходитись окремо від інших підрозділів у певних філіях, або взагалі знаходитись окремо та працювати самостійно завдяки співпраці з іншими філіями та підрозділами. Це зумовлює необхідність відправляти з певною частотою працівників у відрядження. Також відрядження можуть бути зумовлені співпрацею з іншими підприємствами.

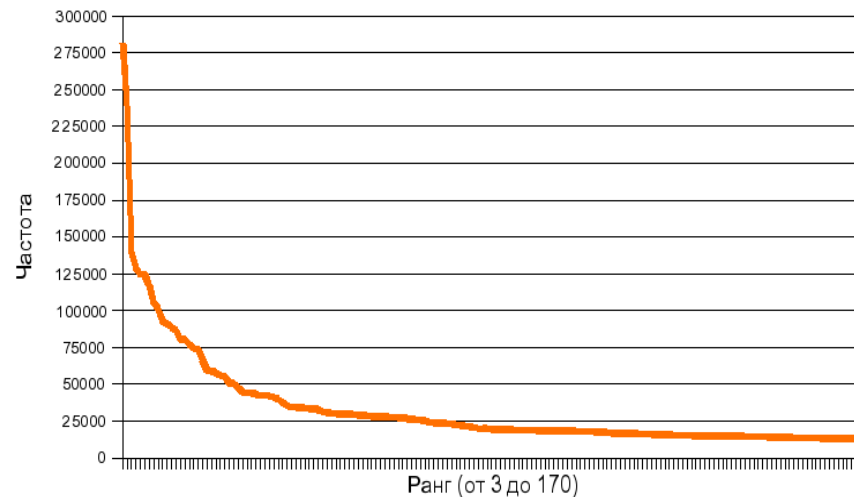
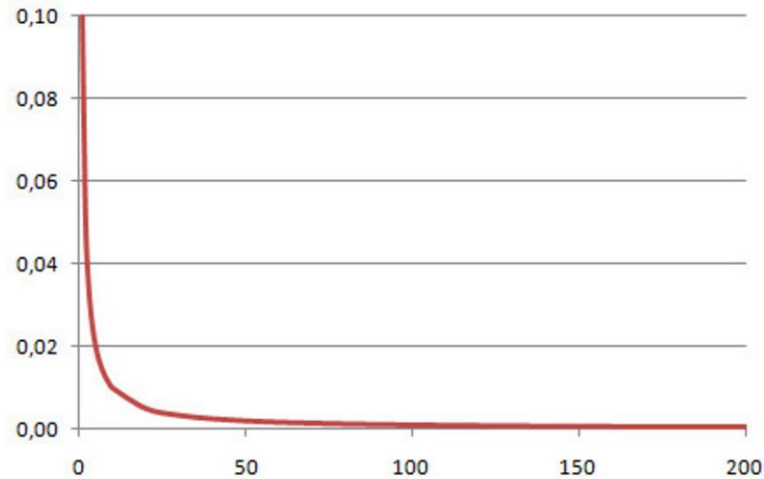
# Аналіз методів та засобів ідентифікації особи



# Закон Ціпфа

емпіричне правило розподілу частоти слів природної мови: якщо все слова мови (або просто досить довгого тексту) впорядкувати по спадаючій частоти їх

використання, то частота  $n$ -го слова в такому списку опиниться приблизно обернено пропорційній його порядковому номеру  $n$  (так званому рангу цього слова, см. шкала порядку)



# Алгоритм «TF-IDF»

TF-IDF (від англ. TF - term frequency, IDF - inverse document frequency) - статистична міра, яка використовується для оцінки важливості слова в контексті документа, що є частиною колекції документів. Вага деякого слова пропорційний кількості вживання цього слова в документі, і обернено пропорційна частоті вживання слова в інших документах колекції.

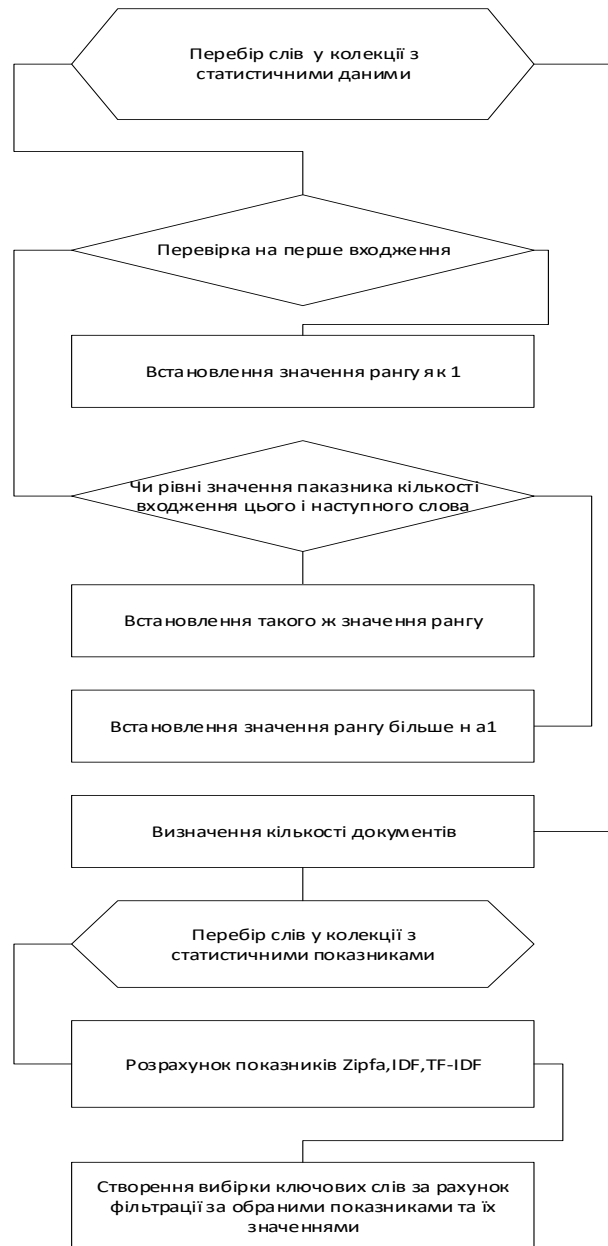
# Функції обрахунку даних

$$Q = \frac{T_f}{T_a}$$

$$Z = \frac{Q * R}{t_a}$$

$$IDF = \log_{10} \frac{D_a}{D_w}$$

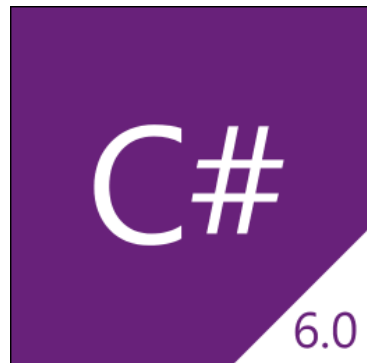
$$TF - IDF = Q * IDF$$





# Вибір технологій

- C#
- Система управління базами даних MS SQL
- Програмна основа ASP.NET Framework



Visual Studio



Microsoft



Microsoft®  
SQL Server®

# Зовнішній вигляд програмного продукту

## Personal documents list

[Back to details](#)

Name	IsPersonally	DateCreate	
Doc1	<input checked="" type="checkbox"/>	1/8/2018 3:47:56 PM	<a href="#">Details</a>
Doc2	<input checked="" type="checkbox"/>	1/8/2018 3:48:05 PM	<a href="#">Details</a>
Doc3	<input checked="" type="checkbox"/>	1/8/2018 3:48:12 PM	<a href="#">Details</a>
Doc4	<input checked="" type="checkbox"/>	1/8/2018 3:48:21 PM	<a href="#">Details</a>
Doc5	<input checked="" type="checkbox"/>	1/8/2018 3:48:32 PM	<a href="#">Details</a>
Doc6	<input checked="" type="checkbox"/>	1/8/2018 3:48:38 PM	<a href="#">Details</a>

## Stats(1160,195,32)

[Back to Details](#)

Word	Amount	Rank	Tf	Zipfa	Idf	Tfidf	IsRemote	DocAmount
disk	12	5	1.034	0.431	0.18	0.186	<input type="checkbox"/>	5/6
engine	11	6	0.948	0.517	0.18	0.171	<input type="checkbox"/>	5/6
event	11	6	0.948	0.517	0.18	0.171	<input type="checkbox"/>	5/6
moody	11	6	0.948	0.517	0.18	0.171	<input type="checkbox"/>	5/6
salary	10	7	0.862	0.603	0.18	0.155	<input type="checkbox"/>	5/6
former	10	7	0.862	0.603	0	0	<input type="checkbox"/>	6/6
accomplish	10	7	0.862	0.603	0	0	<input type="checkbox"/>	6/6
jumper	10	7	0.862	0.603	0.18	0.155	<input type="checkbox"/>	5/6
attitude	10	7	0.862	0.603	0.18	0.155	<input type="checkbox"/>	5/6
precision	10	7	0.862	0.603	0.18	0.155	<input type="checkbox"/>	5/6
mature	10	7	0.862	0.603	0.18	0.155	<input type="checkbox"/>	5/6

# Зовнішній вигляд програмного продукту

## Remoute documents list

[Back to details](#)

Name	IsPersonally	DateCreate		
Doc7	<input type="checkbox"/>	1/8/2018 3:50:06 PM	<a href="#">Details</a>	<a href="#">Differents</a>
Doc8	<input type="checkbox"/>	1/8/2018 3:50:18 PM	<a href="#">Details</a>	<a href="#">Differents</a>
Doc9	<input type="checkbox"/>	1/8/2018 3:50:29 PM	<a href="#">Details</a>	<a href="#">Differents</a>
Doc10	<input type="checkbox"/>	1/8/2018 3:50:39 PM	<a href="#">Details</a>	<a href="#">Differents</a>
Doc11	<input type="checkbox"/>	1/8/2018 3:50:54 PM	<a href="#">Details</a>	<a href="#">Differents</a>
Doc12	<input type="checkbox"/>	1/8/2018 3:51:07 PM	<a href="#">Details</a>	<a href="#">Differents</a>

## Document stats

[Back to document list](#)

### Total

Personal: 1160,195,32

Remoute: 1142,193,30

Document: 141,185

advantage	1	0.541	0.776	0.788	5/6	5/6
increase	2	1.081	0.776	0.788	5/6	5/6
reflect	2	1.081	0.776	0.788	6/6	6/6
event	0	0	0.948	0.963	5/6	5/6
moody	0	0	0.948	0.963	5/6	5/6
degree	0	0	0.862	0.876	5/6	5/6
ccrboy	0	0	0.776	0	6/6	0
ccrb	0	0	0.776	0	6/6	0
attitude	2	1.081	0.862	0.876	5/6	5/6
weak	1	0.541	0.862	0.876	6/6	6/6
launch	2	1.081	0.776	0.788	6/6	6/6
stubborn	1	0.541	0.776	0.788	5/6	5/6
implement	1	0.541	0.776	0.788	6/6	6/6

## Результат виконаної роботи

В результаті роботи удосконалено метод аналізу частотних характеристик лексики та розроблено алгоритм виявлення авторства.

Розроблена комп'ютерна програма надає змогу спростувати або підтвердити авторство, що підвищує інформаційну безпеку.

Дякую за увагу