

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
Черкаський національний університет  
імені Богдана Хмельницького  
Черкаський інститут банківської справи  
Чорноморський державний університет імені Петра Могили

*Всеукраїнська науково-практична  
Інтернет-конференція*

**Автоматизація та комп'ютерно-  
інтегровані технології у  
виробництві та освіті:  
стан, досягнення,  
перспективи розвитку**

*11-21 березня 2021 року*

*м. Черкаси*

2. Джарратано Д. Экспертные системы: принципы разработки и программирование, 4-е издание.: Пер. с англ. - М.: ООО «Вильямс», 2007. - 1152 с.
3. Gruber, T. Toward Principles for the Design of Ontologies Used for Knowledge Sharing // International Journal Human-Computer Studies. - 1995. -Vol. 43. -P. 907-928

*Слюсар В.И., д.т.н., проф.*

*Центральный научно-исследовательский институт вооружения и военной техники Вооруженных Сил Украины, Киев*

## **ЛАТЕНТНО-СЕМАНТИЧЕСКИЙ АНАЛИЗ ДОКУМЕНТОВ НА ОСНОВЕ ТОРЦЕВОГО ПРОИЗВЕДЕНИЯ МАТРИЦ**

Как известно, в теории обработки естественного языка латентно-семантический анализ (LSA) набора документов осуществляется на основе формирования терм-документной матрицы (term-document matrix (TDM), document-term matrix) и её последующего сингулярного SVD-разложения. Терм-документная матрица, по сути, является обобщением матрицы инцидентности, используемой при анализе корпуса текста [1]. При этом отличие TDM состоит в том, что ее строки соответствуют не отдельным предложениям текста, а документам, тогда как элементы столбцов указывают, сколько раз тот или иной термин (слово) встречается в этих документах.

Целью доклада является развитие возможностей LSA на основе торцевого произведения матриц [2 - 4]. Рассмотрим в качестве примера набор из двух документов:

D1 = “я люблю кофе”; D2 = “я не люблю кофе”.

Структуру данного комплекта документов можно представить как “мешок слов” (Bag-of-Words, BoW) в виде табл. 1. Данному BoW соответствует терм-документная матрица вида:

$$G = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

В рассмотренном примере элементы TDM матрицы содержат 1 и 0. В более общем случае её элементы могут быть больше 1.

Таблица 1

	я	люблю	не люблю	кофе
--	---	-------	----------	------

D1	1	1	0	1
D2	1	0	1	1

Для расширенного анализа блока документов предлагается использовать кратные торцевые произведения терм-документной матрицы. В случае торцевого перемножения двух матриц G получим:

$$C = G \square G = \begin{bmatrix} 1 & 1 & 0 & 1 & | & 1 & 1 & 0 & 1 & | & 0 & 0 & 0 & 0 & | & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & | & 0 & 0 & 0 & 0 & | & 1 & 0 & 1 & 1 & | & 1 & 0 & 1 & 1 \end{bmatrix}$$

Соответствующий модифицированный “мешок слов” может быть представлен в виде табл. 2, поясняющей суть элементов результирующей блочной матрицы C.

Таблица 2

я				люблю			
я	люблю	не люблю	кофе	я	люблю	не люблю	кофе
1	1	0	1	1	1	0	1
1	0	1	1	0	0	0	0
не люблю				кофе			
я	люблю	не люблю	кофе	я	люблю	не люблю	кофе
0	0	0	0	1	1	0	1
1	0	1	1	1	0	1	1

Как видно, переход к торцевому произведению TDM позволяет получить информацию о частоте встречаемости тех или иных словосочетаний в анализируемых документах. Например, как следует из табл. 2, во втором документе отсутствует комбинация слов “я люблю”, “люблю кофе” (нулевые элементы). В этом случае не учитывается порядок следования слов, а также их повторяемость. В частности, для обоих документов отмечено наличие двойной комбинации слова “я”, хотя оно использовано в них однократно.

Рассмотренный подход может быть распространён на любой уровень кратности K торцевых произведений, что позволяет охватить латентно-семантическим анализом словосочетания из K слов. При этом SVD-разложение модифицированных TDM позволяет идентифицировать в документах наиболее часто встречающиеся словосочетания. Для расширения возможностей метода по количеству охватываемых слов можно, аналогично [1], использовать квадратичные формы торцевых произведений TDM.

### Список использованных источников

1. Слюсар В.И. Применение торцевого произведения матриц в задачах обработки естественного языка. //Нейромережні технології та їх застосування НМТІЗ-2020: збірник наукових праць XIX Міжнародної наукової конференції «Нейромережні технології та їх застосування НМТІЗ-2020». - Краматорськ: Донбаська державна машинобудівна академія. - 2020. - С. 156 - 162. - DOI: 10.13140/RG.2.2.31568.53762.
2. Слюсар В.И. Торцевые произведения матриц в радиолокационных приложениях// Известия высших учебных заведений. Радиоэлектроника.- 1998. - Том 41, № 3.- С. 71 - 75.
3. Слюсар В.И. Семейство торцевых произведений матриц и его свойства// Кибернетика и системный анализ. – 1999.- Том 35; № 3.- С. 379-384.- DOI: 10.1007/BF02733426
4. Основы военно-технических исследований. Теория и приложения. Том. 2. Синтез средств информационного обеспечения вооружения и военной техники. / А.И. Миночкин, В.И. Рудаков, В.И. Слюсар. – Киев:«Грамна, 2012. – С. 7 – 98; 354 – 521.

*Шевченко Наталя Юріївна, к.е.н., доцент,  
Кравець Анна Ігорівна, студентка бакалавріату  
Донбаська державна машинобудівна академія,  
Краматорськ*

## **АНАЛІЗ ГОЛОСОВИХ ПОВІДОМЛЕНЬ ПРИ ВИЗНАЧЕННІ ІНДИВІДУАЛЬНИХ ХАРАКТЕРИСТИК КОРИСТУВАЧА МОБІЛЬНОГО ПРИСТРОЮ**

Одним з ключових біометричних параметрів людини є її голос, що має набір індивідуальних особливостей, які відносно легко піддаються вимірюванню, наприклад, частотні або амплітудні характеристики голосового сигналу [1].

Серед базових характеристик голосу можна виділити: силу голосу або гучність;

висоту голосу (у середньостатистичної людини діапазон до 1,5 октави, в повсякденному спілкуванні більшість використовує 3-4 ноти);

тембр голосу – сукупність додаткових коливань або обертонів, які виникають поряд з основною частотою;

тон голосу (емоційне забарвлення голосу), яке задається різними методами: зміною гучності, темпу, висоти.

Алексеева Г.М., Кравченко Н.В., Горбатюк Л.В., Черкательська Г.О. ТЕХНОЛОГІЇ – ЯК ЗАСІБ САМОВДОСКОНАЛЕННЯ: НАВИЧКИ SOFT SKILLS .....	238
Алексеева Г.М., Данилова Н.О. ТЕХНОЛОГІЇ – ЯК НЕОБХІДНА СКЛАДОВА ДЛЯ ФОРМУВАННЯ SOFT SKIL .....	240
Білик Валерія, Алексеева Г.М. ВИКОРИСТАННЯ PREZI ДЛЯ РОЗРОБКИ ВІДЕО ЛЕКЦІЙ .....	243
<b>Секція 7. Проблеми підготовки фахівців у галузі автоматизації та інформаційних технологій .....</b>	<b>246</b>
<i>Ткаченко А.В.</i> ОРГАНІЗАЦІЯ ТЕСТОВОГО КОНТРОЛЮ ЗНАНЬ З ФІЗИКИ.....	247
<i>Psol Serhij Vasylyovych, Rudyk Oleksandr Yuhymovych</i> METHODS OF REMOTE FORM OF KNOWLEDGE CONTROL .....	251
<i>Верецзак І.А., Наконечна О.А.</i> ОСНОВНІ ВІДМІННОСТІ МІЖ WEBSITE I WEB APPLICATION .....	253
<i>Кравченко В.И., Стукалова Ю.А.</i> ПРОБЛЕМЫ ПОДГОТОВКИ В МАШИНОСТРОИТЕЛЬНОМ ВУЗЕ СПЕЦИАЛИСТОВ ПО АВТОМАТИЗАЦИИ РЕШЕНИЯ БИОИНЖЕНЕРНЫХ ЗАДАЧ.....	256
<i>Ішутіна О.Є.</i> ПІДГОТОВКА МАЙБУТНІХ УЧИТЕЛІВ ПОЧАТКОВОЇ ШКОЛИ ДО ВИКОРИСТАННЯ AR-ТЕХНОЛОГІЙ .....	258
<b>Секція 8. Інтелектуальні системи та машинне навчання .....</b>	<b>261</b>
<i>Petro Shepita</i> METHODS OF ANALYSIS AND CLASSIFICATION OF GRAPHIC INFORMATION IN INTELLECTUAL SYSTEMS .....	262
<i>Шеніта П.І.</i> ЗАСТОСУВАННЯ ФРЕЙМОВОЇ МОДЕЛЬ ПРЕДСТАВЛЕННЯ ЗНАНЬ ПРИ ПРОЕКТУВАННІ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ УПРАВЛІННЯ ПОЛІГРАФІЧНИМ ПІДПРИЄМСТВОМ .....	264
<i>Слюсар В.И.</i> ЛАТЕНТНО-СЕМАНТИЧЕСКИЙ АНАЛИЗ ДОКУМЕНТОВ НА ОСНОВЕ ТОРЦЕВОГО ПРОИЗВЕДЕНИЯ МАТРИЦ.....	266
<i>Шевченко Н.Ю., Кравець А.І.</i> АНАЛІЗ ГОЛОСОВИХ ПОВІДОМЛЕНЬ ПРИ ВИЗНАЧЕННІ ІНДИВІДУАЛЬНИХ ХАРАКТЕРИСТИК КОРИСТУВАЧА МОБІЛЬНОГО ПРИСТРОЮ .....	268