

Ministry of Education and Science of Ukraine
Sumy State University (Ukraine)
The College of New Jersey (USA)
Camarines Sur Polytechnic Colleges (Philippines)
European Danube Academy (Germany)
University of Kragujevac (Serbia)



Proceedings

STRATEGIC INNOVATIONS OF SOCIAL COMMUNICATIONS AND FOREIGN PHILOLOGY IN CRISIS TIMES

I International Scientific and Practical Conference

1 June 2024

Sumy State University, Ukraine



Sumy 2024

**Ministry of Education and Science of Ukraine
Sumy State University (Ukraine)
The College of New Jersey (USA)
Camarines Sur Polytechnic Colleges (Philippines)
European Danube Academy (Germany)
University of Kragujevac (Serbia)**

Proceedings

**STRATEGIC INNOVATIONS OF SOCIAL COMMUNICATIONS AND
FOREIGN PHILOLOGY IN CRISIS TIMES**

I International Scientific and Practical Conference

1 June 2024
Sumy State University, Ukraine

Sumy 2024

UDC: 81'25'42:316.7(063)(0.034)

S88 Strategic Innovations of Social Communications and Foreign Philology in Crisis Times: collection of scientific paper of the I International Scientific and Practical Conference (1 June 2024) / responsible editor Nabok M., responsible designer Sadivnycha M. Sumy: Sumy State University, 2024. 351 p.

The collection includes scientific materials devoted to current problems in the field of distance teaching and learning in the era of digital technologies, in particular in language pedagogy, media innovations, modern translation, ethnolinguistics, stylistics, methods and methodologies of modern linguistic and literary disciplines. The composition of the collection is scientific research on the preservation of national identity and cultural heritage, intercultural communications, psychological theory and practice, sociological research.

For teachers of higher education institutions, graduate students, students, teachers of secondary schools, gymnasiums, lyceums and colleges, all those interested in socio-humanitarian issues.

The materials are printed in the author's editorial

Sumy State University, 2024

Brands Through the Power of Archetypes. New York: McGraw Hill.

Бодріяр, Ж. (2004). Симулякри і симуляція. Київ: Видавництво Соломії Павличко «Основи».

Федорієнко В.А.

ORCID: <https://orcid.org/0000-0002-0921-3390>

кандидат технічних наук

начальник науково-дослідного відділу інституту стратегічних комунікацій

м. Київ, Україна

Прокопенко О.С.

ORCID: <https://orcid.org/0000-0002-5482-0317>

доктор філософії

начальник науково-дослідної лабораторії

Національного університету оборони України

м. Київ, Україна

Кульчицький О.С.

ORCID: <https://orcid.org/0000-0002-4901-0192>

начальник науково-дослідної лабораторії

Національного університету оборони України

м. Київ, Україна

ОСОБЛИВОСТІ ВИКОРИСТАННЯ ТЕХНОЛОГІЙ ЛІНГВІСТИЧНОГО МОНІТОРИНГУ ІНФОРМАЦІЙНОГО ПРОСТОРУ

Сьогодні, групи методів Data Science та аналітики текстових даних, є найпоширенішими, і вони використовуються для обробки природної мови зокрема, для проведення моніторингу, аналізу та прогнозування змін в інформаційному просторі. Практичне застосування цих технологій полягає у вирішенні задач в сфері стратегічних комунікацій, зокрема для автоматизації процесів моніторингу інформаційного простору, що сприяє підтримці прийняття рішень. Очевидно, для цього важливими є пошук закономірностей певних подій

та явищ в інформаційному просторі, які стають більш доступними лише після залучення до аналізу великого масиву текстових та медіа даних, які є доступними в Інтернеті. Визначення штучного походження «всплесків» публікацій може свідчити про прояви інформаційно-психологічних операцій з боку рф, які мають на меті викликати негативні емоції у населення України. Для такого пошуку використовуються програмні веб платформи моніторингу інформаційного простору, які стали набувати все більшого поширення з початку повномасштабного вторгнення рф на територію України. Такі інструменти використовуються, для швидкої візуалізації динаміки змін за рахунок аналізу текстових повідомлень, у тому числі із соціальних мереж (Прокопенко, 2023). Серед них такі веб платформи як, YouScan, Attack Index, Brand24, Looqme, SemanticForce тощо. Адже використання ефективних механізмів пошуку російського впливу в інформаційному просторі за ключовими словами на основі методів машинного навчання у текстових повідомленнях є актуальним завданням.

Загалом, в інформаційному просторі міститься інформація у різних об'ємах, форматах тощо. З цієї інформації можна відшукати російські наративи направлені проти України, які стали особливо агресивними від початку повномасштабного вторгнення. Цільовими аудиторіями розповсюдження російських інформаційно-психологічних продуктів стали аудиторії: українська, російська, країн члені НАТО. Для ефективного пошуку наративів у подібних системах береться до уваги усі доступні дані, у тому числі, ті, які включають зображення, географічні назви, описи, теги, відомості про самі джерела інформації тощо. Досить широкого розповсюдження набув метод комп'ютерного зору, який допомагає у вилученні текстів із зображень. Також, можна мати уявлення і про особливості цільові аудиторії шляхом аналізу коментарів у соціальних мережах та поширенні відео. Зібрана під час моніторингу та якісно проаналізована інформація необхідна для прийняття рішення на верхньому рівні.

Ранжування або сортування інформації у каталозі тем із подальшим

формуванням сюжетів може містити механізми лінгвістичного перекладу контенту на основі автоматичної модерації (NLTK, 2024), що дозволяє фактично здійснювати категоризацію новин чи дописів, визначення дублікати повідомлень та їх штучне їх приховування. Візуальний пошук базується на використанні обробки тексту і містить доволі великий набір технологій та засобів для роботи з ними з подальшим моделюванням цих даних і представленням певних тенденцій розвитку.

Варто зазначити, що обробка природної мови базуються на роботі з текстовими документами, що є базовим об'єктом функціонування системи автоматизованого моніторингу. Ці тексти можуть мати довільний вигляд, а в залежності від предметної області вони різняться, також вони не мають накладеного стандартного обмеження на розмір. Тобто, це може бути, як одне слово, так і, наприклад, законодавчий акт на тисячі і десятки тисяч слів. Це фактично інформація записана людиною і є виражена словами, цифрами розділовими знаками тощо. Але зазвичай цікавлять не поодинокі документи, а набори текстових документів, які називаються корпусами. Приклад, такого корпусу з кількох документів містить назви тем, які включають різну інформацію про кількісно-якісні характеристики кожної із теми для такої моделі.

Характеристики можуть містити композитні особливості документів. Їх формування та пошук необхідний для побудови моделі алгоритмів чи правил за якими робиться аналіз і певне передбачення, що може бути необхідним для вирішення певної прикладної чи дослідницької задачі. При цьому, самі текстові документи містять неструктурованою інформацією і цей текст потрібно подробити для того щоб алгоритм міг сприйняти це на вхід у формі певних числових векторів. Але перш, ніж перетворити у числове представлення різнорідних текстів варто зосередити увагу на їх «токенізації» (це процес який ділить один суцільний рядок на набір, так званих, атомарних будівельних блоків, які називаються токени). У залежності від задачі і специфіки алгоритму - це можуть бути різні структури. В даному випадку, це просто слова розділені за пробілом, у яких зберігся, в тому числі, в складі їхніх реєстрів, їх форма, яка

може змінюватися. Це пояснюється тим, що токенами можуть бути як окремо взяті символи, так, і склади, де можуть враховуватися реєстри, а в певних задачах – формуватися звуження словника унікальних слів. Токени далі схематизуються, тобто приводяться до певної нормальної форми, де іменники переводяться в однину та називний відмінок, дієслова в інфінітив (якщо це можливо). Після цього, такий текст стає менш зашумленим і наслідуює менше різноманітних форм одного і того самого терміну. Це досягається за допомогою використання певних спеціальних словників і структур даних. Але така техніка включає набір технік, і зокрема, алгоритми прогнозування, що потребують числове введення.

Якщо набір слів впорядкований, то він стає більш зрозумілим, а наступним етапом стає використання способів конвертації. Вони передбачають адаптацію до більшості унікальних слів і формування потрібних корпусів. Далі будується впорядкований словник для кожного документу в заданому корпусі. Через цей словник пропускаємо і отримуємо числові вектори фіксованої розмірності. Тобто, вектори визначаються двійковими значеннями (без слів – нулі, наявність слів – одиниця). За кількістю появи слів, які зустрічаються у такому документі, виходить векторне представлення має досить розріджений вид. З таким алгоритмом працювати простіше зокрема, використовувати його для прогнозування на основі, наприклад, логістичні регресії. Адже, якщо за допомогою цієї моделі можна зробити класифікацію текстових даних за визначенням якоїсь категорії, і для кожного документу є змога його «натренувати» або навчити автоматичним змінам. Це можливо досягти шляхом оптимізації ваги цієї лінійної моделі на наборі розріджених векторів. Після цього, отримується алгоритм, який має на вхід такий же розріджений вектор, що відповідає нормальному розподілу ймовірностей. Для цього переліку категорій можливо застосувати принцип найбільшої правдоподібності, щоб визначити, якій саме категорії відповідає обраний текстовий документ. У даному прикладі найвища ймовірність вважається переможем у цьому змаганні. А на практиці це вчасне реагування на зміни в інформаційному просторі та протидія ворожому

інформаційному впливу, за рахунок визначення найбільш чітких тем, які відповідають російським наративам.

Список використаних джерел:

Documentation. Natural Language Toolkit (2024). <https://www.nltk.org/>

Прокопенко, О., Федорієнко, В., Кульчицький, О. (2023). Підхід щодо виявлення і аналізу інформаційних загроз національній безпеці України у системі стратегічних комунікацій. *Збірник наукових праць Центру воєнно-стратегічних досліджень Національного університету оборони України, № 2 (73)*. 35–43. <https://doi.org/10.33099/2304-2745/2023-2-78/35-43>