

**Пономарьов О.В.** студент гр. 124-20м-1

**Науковий керівник: Коряшкіна Л.С., к.ф.-м.н., доцент кафедри системного аналізу і управління**

*(Національний технічний університет "Дніпровська політехніка", м. Дніпро, Україна)*

## **ФІЛЬТРАЦІЯ ОСНОВНИХ МОМЕНТІВ ЗАПИСІВ ОНЛАЙН-ТРАНСЛЯЦІЙ НА ОСНОВІ АНАЛІЗУ КОМЕНТАРІВ**

Більшість досліджень у галузі обробки природної мови зосереджено на моделюванні та розумінні тексту, що складається з повних речень з правильним написанням та граматикую. Однак чат у прямому ефірі кардинально відрізняється від звичайного тексту. Глядачі зазвичай пишуть короткі повідомлення, відповідаючи на події в трансляції, часто з неправильною граматикую і безліччю сленгових виразів, що повторюються (слова, смайлики). Крім того, сленгові вирази, які зазвичай використовуються в чаті у прямому ефірі, невідомі традиційним засобам щодо розуміння мови, які зосереджені на прозовому тексті. Оскільки прямі трансляції стають все більш популярними, стає все більш актуальним необхідність вивчати за допомогою аналізу чату способи взаємодії аудиторії з трансляцією. Це непросте завдання, чат у прямому ефірі - це багата і складна область. Чат трансляції містить безліч орфографічних помилок, наприклад через те, що глядачі швидко набирають повідомлення, які реагують на потік. Ці орфографічні помилки можуть бути навмисними. Крім того, у чатах широко використовуються «емоції», специфічні для предметної області смайли з багатим та складним значенням. Хоча чат у прямому ефірі має деякі подібності з іншими соціальними медіа даними, у літературі чітко показано, що чат у прямому ефірі має кілька унікальних властивостей. Насамперед, він тісно пов'язаний з потоковим ігровим контентом, наприклад тому, що учасники реагують на події в потоці, і, таким чином, чат відіграє ключову роль у розумінні контексту прямих трансляцій. По-друге, величезний масштаб розміру чату в доповненні до часових обмежень призводить до унікальних властивостей. Таким чином, розуміння цієї галузі є одночасно надзвичайно складним і дуже важливим, оскільки популярність потокової передачі зростає як розважальна область.

Метою даної роботи є аналіз коментарів під онлайн-трансляціями і фільтрація ключових моментів останніх задля можливості повторного перегляду лише цих найцікавіших (за реакцією глядачів) відрізків трансляцій. Запропонований наступний підхід щодо вирішення проблеми виявлення основних моментів. На вхід системи подаються коментарі в одиницю часу (секунди). По тому, наскільки активно пишуть в момент часу користувачі трансляції (реагуючи на той чи інший побачений ними момент, але так само це може бути звичайна активність, яка викликала інтерес, наприклад: відповідь на питання від ведучого трансляції глядачам і таке інше), визначаються відрізки за часом, які є ключовими протягом всієї трансляції (наприклад, щоб не дивитися весь матч, можна переглянути лише ключові моменти). На даний момент виявлення кращого моменту відбувається за рахунок кількості коментарів за одиницю часу, коли багато людей починають щось писати, як тільки побачили щось надзвичайне або заради чого вони дивилися трансляцію дії, які змогли вивести команду, за яку хворієш, наперед до перемоги, якщо проводити аналогію з футболом, то це як крики з трибун після забитого м'яча у ворота суперника. У подальшому планується пошук слів у повідомленнях, що найчастіше зустрічаються, і їх аналіз (здебільшого це смайлики або сленгові слова) з метою подальшого порівняння і класифікації декількох різних трансляцій.