

РЕАЛІЗАЦІЯ ЗАДАЧ АВТОМАТИЗОВАНОГО ОЦІНЮВАННЯ ОПІНІЇ ТЕКСТОВОЇ ІНФОРМАЦІЇ У WEB-ДОКУМЕНТАХ

Доповідь присвячена прикладним аспектам реалізації задач автоматизованого оцінювання опінії текстової інформації [1]. У доповіді розглянуто структурну схему процедури добування опінії текстової інформації з web-документів.

Задачі автоматизованого оцінювання опінії текстової інформації поділяють на три основні категорії: оцінювання на рівні документів, на рівні речень та на рівні ознак.

Добування опінії текстової інформації проходить в декілька етапів. Маючи web-документ, потрібно виокремити з нього коментарі споживачів, що забезпечується html-парсером, котрий видаляє html теги з web-документу та формує чистий текст коментаря. Після виокремлення, коментар проходить через POS-парсер (POS – part of speech tagging), який розподіляє речення на частини мови. POS-парсер формує XML-документ, в якому кожному слову, відповідно до частини мови, присвоюється окремий тег. Аналіз контексту речення відбувається на основі бази даних, в котрій прописані слова з їх позитивними та негативними значеннями.



Рис 1. Структурна схема процедури добування опінії текстової інформації з web-документів

Щодо добування опінії на рівні документів – ця задача реалізовується як наслідок добування та оцінювання опінії на рівні речень, шляхом зведення інтегрального показника оцінених речень.

У доповіді детільніше звернено увагу на задачі оцінювання опінії текстової інформації на рівні речень. На цьому етапі потрібно ідентифікувати суб'єктивне речення та класифікувати його емоційне забарвлення. Суб'єктивним реченням вважається таке речення, що містить опінію про даний об'єкт.

На даний час реалізовано html-парсер та POS-парсер, ведеться робота над аналізом контексту словосполучень. Добування опінії текстової інформації на рівні ознак [2] варто реалізовувати після того, як буде проведено аналіз контексту словосполучень. Вивід результату здійснюється у вигляді порівняльних діаграм та у вигляді загальної інтегральної оцінки опінії.

Література.

1. Хомів Б. А. Задачі оцінювання опінії текстової інформації у web-документах / Б.А. Хомів, С.А. Лупенко, О.А. Пастух / Матеріали міжнародної науково-практичної конференції «Фундаментальні та прикладні проблеми сучасних технологій», Тернопіль, 19 – 21 трав. 2010 р. – Тернопіль: ТНТУ. – С. 324-325.

2. Liu B. Opinion Observer: Analyzing and Comparing Opinions on the Web / B. Liu, M. Hu, J. Cheng / To appear in Proceedings of the 14th international World Wide Web conference (WWW-2005), Japan, May 10-14, 2005, pages 342-351.