

РЕФЕРАТ

Дипломна робота: X с., 2 табл., рис., 2 дод., 21 джерело

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, ІНФОРМАЦІЙНИЙ ПОШУК, ТЕКСТОВА АНАЛІТИКА, МЕДИЦИНА, ОБРОБКА ПРИРОДНОЇ МОВИ, КЛАСТЕРИЗАЦІЯ, ХМАРНІ ОБЧИСЛЕННЯ.

Актуальність роботи: на сьогодні більша частина інформації знаходиться у неструктурованому вигляді, тому використання її звичними аналітичними моделями являється неможливим. В медичній сфері, обробляючи текстові дані, можливо значно покращити якість зворотнього зв'язку з пацієнтами, а тому і якість препаратів.

Об'єкт дослідження: медичні текстові звіти, а саме – відгуки пацієнтів.

Предмет дослідження: інформаційний пошук, інтелектуальний аналіз, метод максимальної правдоподібності, латентно-семантичний аналіз, булеві правила.

Мета роботи: дослідження існуючих методів обробки неструктурованих текстових даних та їх впровадження у системі аналізу та категоризації текстової медичної звітності.

Метод дослідження: використання знань про обробку природної мови (NLP), математичних методів та моделей для класифікації та кластеризації текстової інформації.

Результати роботи: проведено аналіз сучасних методів інформаційного пошуку, досліджені можливості використання наявних інструментів текстової аналітики, налаштування їх під конкретну галузь, розроблена система, яка дозволяє створювати категорії з певним ступенем вірогідності виконувати задачу класифікації, отримуючи на вхід велику кількість медичних звітів.

Шляхи подальшого розвитку предмета дослідження - удосконалення обраної архітектури моделі, розширення колекції текстів, розмітка даних експертами.