

АНОТАЦІЯ

Об'єктом дослідження являються великі об'єми текстових даних різної природи .

Метою роботи є розроблення інформаційно-категоризаційного модулю категоризації неструктурованих даних. У роботі проаналізовані існуючі методи категоризації неструктурованих даних, визначено проблематику думок та їх вираження, особливості обробки різних точок зору, виявлення спаму.

В роботі розглянуто і проаналізовано «Національний класифікатор України надзвичайних ситуацій». Наукова новизна отриманих результатів полягає в реалізації інформаційно-категоризаційного модулю категоризації неструктурованих даних та реалізації правил та логіки категоризації для класифікатора надзвичайних ситуацій.

Реалізовано алгоритм перетворення описів кодів державного класифікатора у ієрархію значущих фраз, алгоритм очистки ієрархії згідно правил побудови категоризатора.

Практичними результатами роботи являються правила категоризації неструктурованих даних та модуль категоризації неструктурованих даних.

В подальшому рекомендується проводити дослідження в таких напрямках, як автоматизація видобування правил категоризації українських категоризаторів.

**КАТЕГОРИЗАЦІЯ, ТАКСОНОМІЯ, НЕСТРУКТУРОВАНІ ДАНІ,
АНАЛІЗ ЕМОЦІЙНОГО ЗАБАРВЛЕННЯ**