

УДК 004.8

Куликовська Н.А.¹, Азаров В.В.²

¹ асист. каф. КСМ НУ «Запорізька політехніка»

² студ. гр. КНТ-537сп НУ «Запорізька політехніка»

АНАЛІЗ ЯКОСТІ ТЕКСТУ

У мережі Інтернет існує безліч сайтів, наповнених статтями різних якості та спрямованості: одні написані з метою надання користувачам інформації, інші - з метою продати той чи інший товар. Для оцінки подібних статей, необхідно визначити критерії: в рамках поставленого завдання їх зручно представити у вигляді чисел: від 0 до 1, де 0 - найгірший результат за конкретним критерієм. SEO-фахівці, що займаються наповненням сайтів контентом, нерідко перевіряють написані статті за допомогою різних інтернет-сервісів.

Сервіс Google Search Console дозволяє проаналізувати текст на водність, перевірити його орфографію і показати самі часто вживані словосполучення в тексті. Даний сервіс зручний для SEO фахівців, так як він дозволяє зробити перевірку написаних статей простіше. Тут, як і в інших системах, застосовується наступний алгоритм обчислення водності: обчислюється число стоп-слів в тексті, після чого отримане значення ділиться на загальне число слів. Схожа система SEO-CRM., реалізує такий же алгоритм для аналізу тексту, але з використанням максимального і мінімального меж значення оцінки: в даному веб-додатку тексти, більш ніж на 60% складаються з стоп-слів, отримують максимально негативну оцінку, так як вони не несуть в собі ніякої інформації.

В таблиці 1 наведені основні параметри якісного тексту з діапазоном

норми.

Таблиця 1 - Параметри якісного тексту

№	Показник	Діапазон норми, %	На що вказує
1	Унікальність	90-100	Відсутність/наявність копій тексту
2	Нудота	9	Співвідношення повторюваних слів до загального обсягу тексту
3	Водність	40-60	Наявність слів, що не несуть корисної інформації
4	Заспамленість	3-5	Частота вживання ключових слів