ROBERT GORDON UNIVERSITY ABERDEEN

SCHOOL OF COMPUTING SCIENCE AND DIGITAL MEDIA

MSc Big Data Analytics

Student Name:Andrea Perera	Matriculation Number:
Supervisor:Pumudu Fernando	Second Marker:
Project Title: Automatic content quality measurements of technical articles/blogs	
	Start Date:
	Submission Date: August 2022

CONSENT

I agree ☑

I do not agree □

That the University shall be entitled to use any results, materials or other outcomes arising from my project work for the purposes of non-commercial teaching and research, including collaboration.

DECLARATION

I confirm:

Abstract

The use of the internet is growing every day as technology advances. Technical articles/blogs are appealing to readers and researchers due to their ability to express a wide range of opinions and knowledge on a variety of topics and technology trends. As interests in how individuals obtain information changes, research on blog quality has grown in importance. The rapid expansion of this online environment creates a significant need for strengthening the quality of articles/blogs.

In some instances, humans are still involved to assess the quality of article content and it's a time-intensive process and requires more resources. Conventional methods that only adopt page views or article popularity quality indexes to evaluate the quality of an article. Most experts examined how to improve the quality of the article based on SEO in order to place articles in the top search results, rather than the article's content.

The author proposes a system which will focus on evaluating article quality on the article content-based features which is measuring article content breadth and depth along with other features which are the usage of valid URLs, Images, Tables/ diagrams and code and usage of the Expertise / experience Personal opinions by giving a score. Content breadth score is a score/rating of how many related subjects/topics are covered within the article content and Content depth score is a score/rating of how detailed information coverage of a specific topic is within the article. Evaluated the proposed system using human annotated scores for content breadth and depth confusion matric along with accuracy of 70% and 60%.

Keywords: Article; Blog; Article quality; Blog quality; Blogosphere; Quality; Quality measurement; Content quality, Quality assessment, Content based model , Key word Extraction