

Information Retrieval Project Proposal

Group	G19 – Team Thesaurus (150042C, 150234R, 150264H)	
Title	Information Retrieval System for BBC News	
GIT Repo	https://150234R@git.uom.lk/150234R/Information-Retrieval-System-for-BBC-News	
Corpus	BBC Sinhala News [https://www.bbc.com/sinhala]	
Top 10 Queries		
Query No.	Query	Description
1.	title: ක්‍රිකට්	Term Query
2.	title: ශ්‍රී ලංකා	Wild Card Queries
3.	description: "හොංකොං රාජ්‍යය "	Phrase Query
4.	ශ්‍රී ලංකා ක්‍රිකට් ඉදිරි ගමනට යෝජනා පහක්	Stemming - HanSpellStemFilter
5.	විශ්වවිද්‍යාල	Default Search Without Specifying Key Field
6.	{!term f=title}ලංකාවේ	Ranking Results
7.	title: දිශන	Synonyms
8.	title: ශ්‍රී ලංකාවට පළමු ලෝක කුසලාන ජයග්‍රහණය	Stopwords
9.	description: ලිපි මාලිංග	Protected stemming

10.	ഭൂക്കാതലി	Spell Checking (Suggestions are provided)
-----	-----------	---