COVID-19 MLIA · TASK 2 MULTILINGUAL SEMANTIC SEARCH

AN INFORMATION RETRIEVAL SYSTEM FOR COVID-19 DOCUMENTS IN SPANISH

José Alberto Mesa Murgado, Pilar López Úbeda, Manuel Carlos Díaz-Galiano & María-Teresa Martin-Valdivia
Universidad de Jaén

Universidad de Jaén

HELLO !

ABOUT ME

My name is **José Alberto Mesa Murgado** currently working at the University of Jaén, Spain as a researcher for the group SINAI in which we are focused on the study of Human Languages Technologies (HLT)

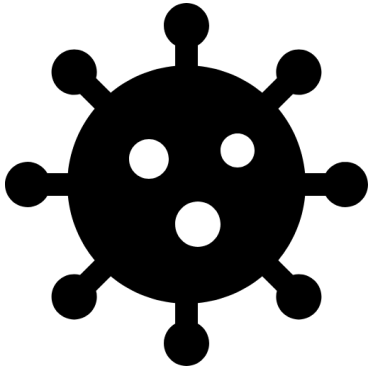
Contact me

@murgado

jmurgado@ujaen.es

Universidad de Jaén

Our participation



Covid-19 has arise a global concern for everyone

A huge number of publications have been made

Provides an interesting challenge for us

Our participation

For our **first entry** to Covid-19 MLIA for which we inscribed in the 9th of november 2020, our initial plan was to deal with both task 1 and task 2 but after considering the time we had available we decided to only focus on the latter and both of its subtasks through a common methodology.

Methodology

Document indexation

With the Corpus in spanish, we decided to use elastic as the framework with which to index every single one of the given documents, this indexation was performed through

Document **title**

Document **keywords**

Document **content**

Run name	Components
sinai1	Search using **keywords**
sinai2	Search using **conversational**
sinai3	Search using **explanation**
sinai4	Search using **keywords, conversational & explanation**
sinai5	Search using **keywords** as query statement on fields: **relevant keywords**

Methodology

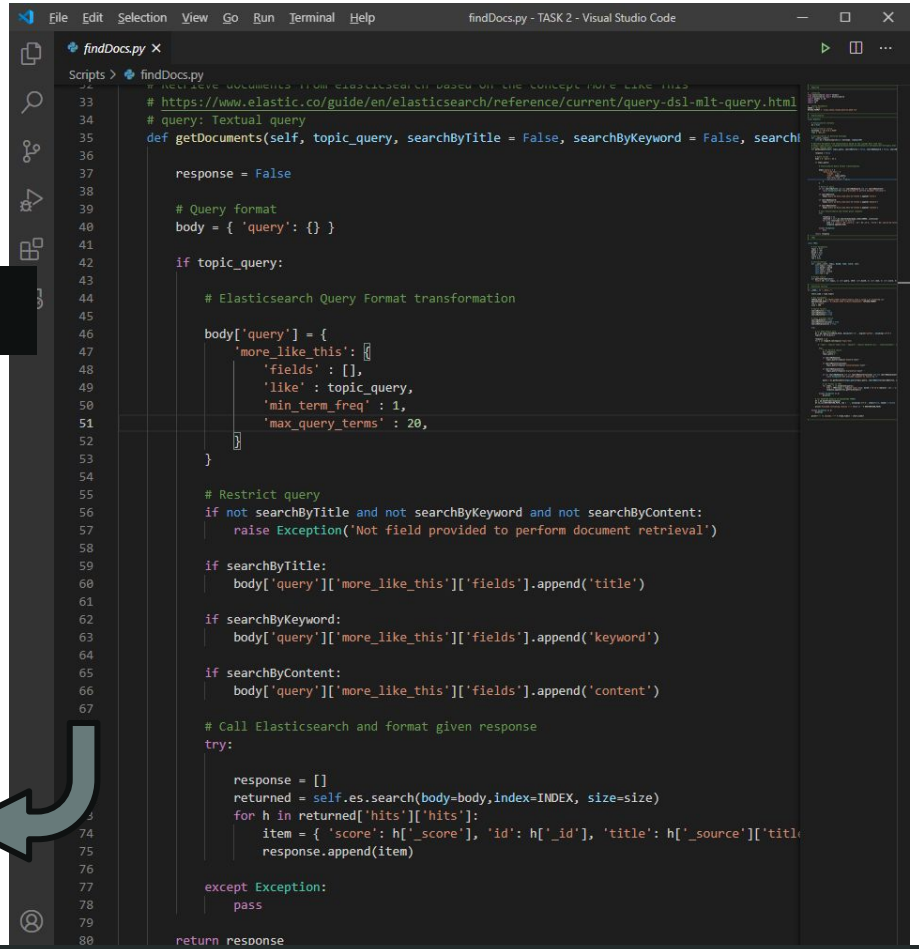
Information retrieval through elastic

Elastic implements Okapi BM25 and uses it by default for its similarity module under the parameters:

b	0.75
k1	1.2

$$\sum_{i}^{n} IDF(q_i) \frac{f(q_i, D) * (k1 + 1)}{f(q_i, D) + k1 * (1 - b + b * \frac{fieldLen}{avgFieldLen})}$$

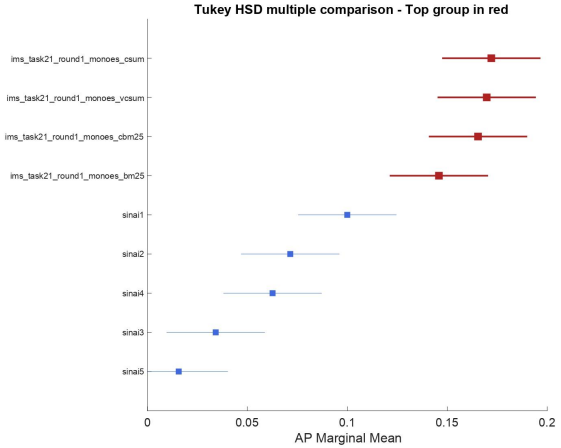
1	Q0	medisys-es-2020_04_3337_93	0	6.3812103	sinai1
1	Q0	medisys-es-2020_03_19723_20	1	6.379256	sinai1
1	Q0	medisys-es-2020_03_2265_191	2	6.36521	sinai1
...	...	...	...	...	..

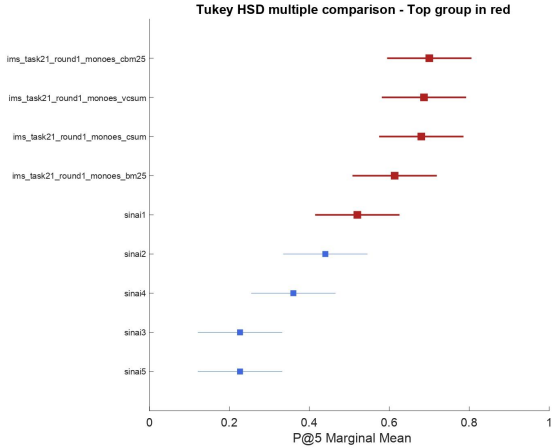


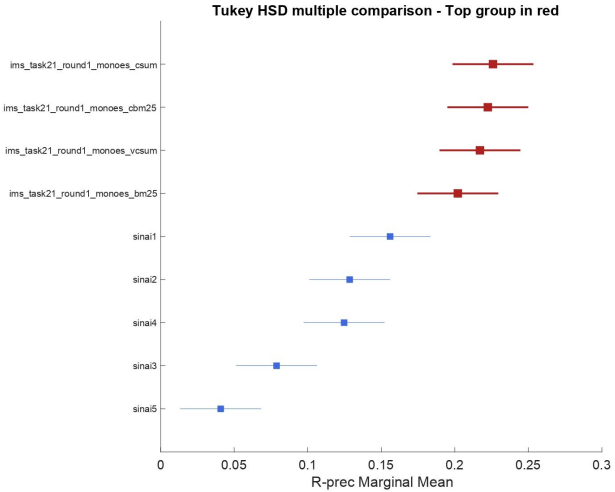
Results I

Task 2 · Subtask 1

High Precision



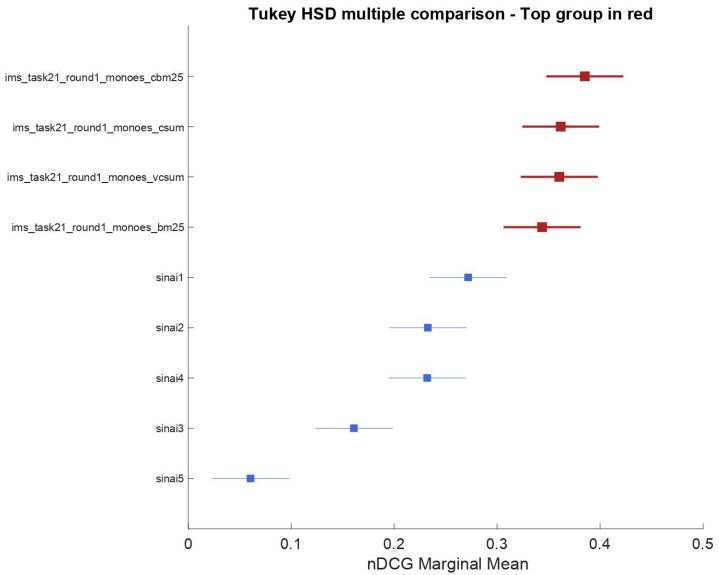


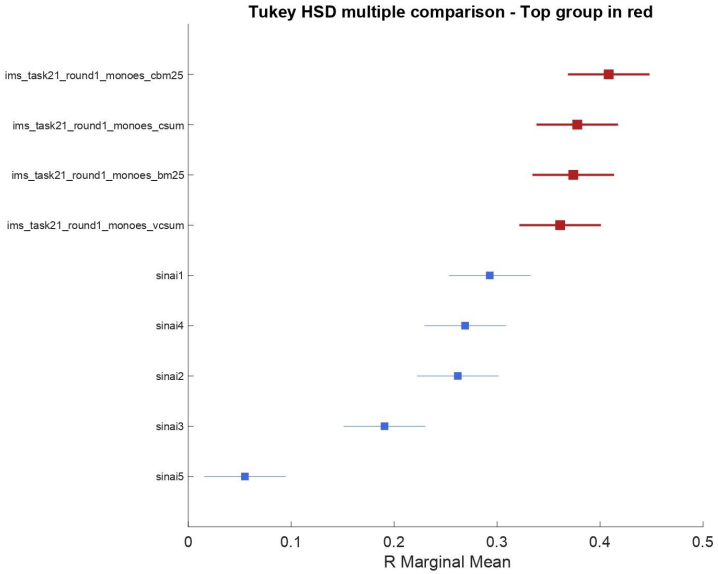


Results I

Task 2 · Subtask 1

High Precision

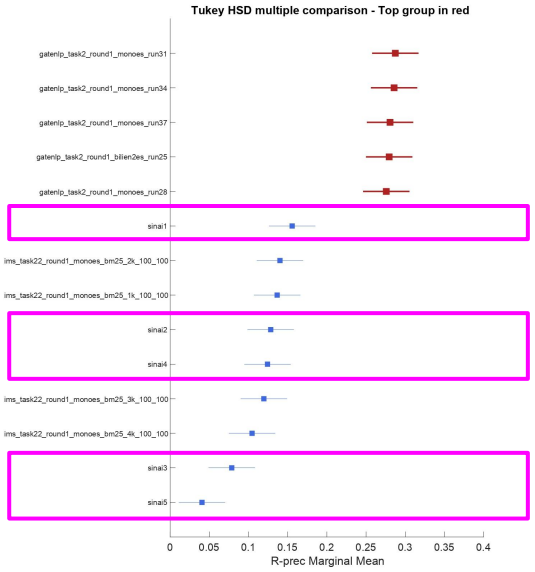


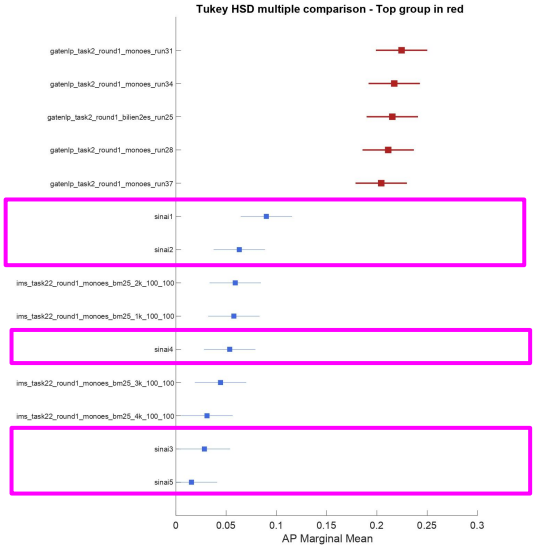


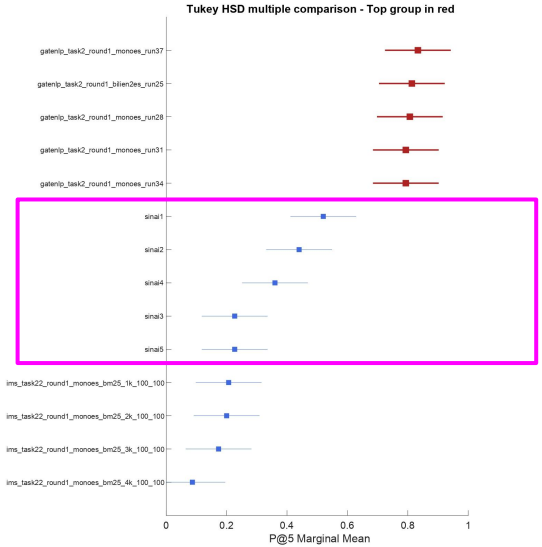
Results II

Task 2 · Subtask 2

High Recall



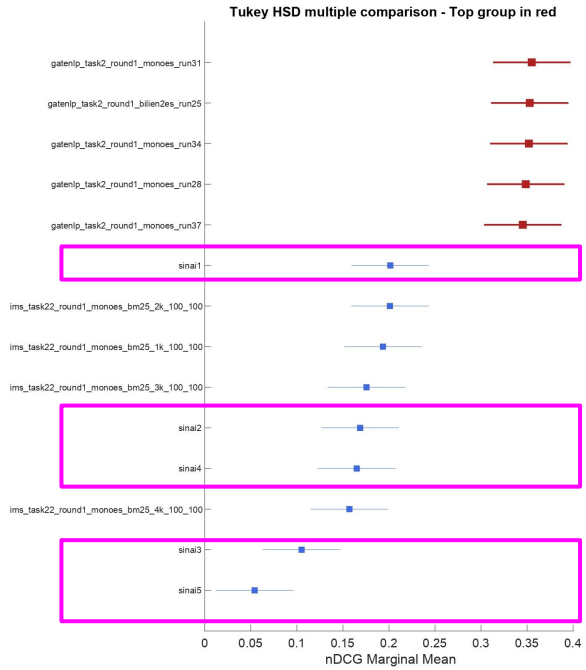


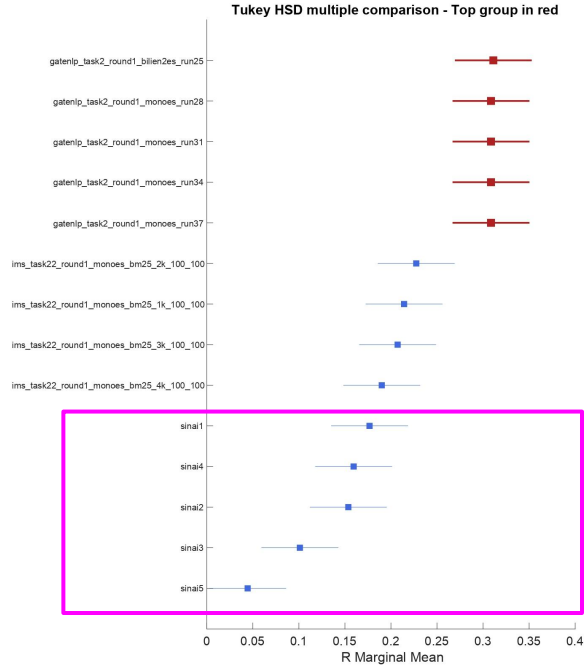


Results II

Task 2 · Subtask 2

High Recall





Discussion & Future work

Discussion

This is a **preliminary approach** for our participation on the initiative which we will improve

Shortage of participants means that no major comparisons can be made

We have to make i**mprovements on our systems in terms of precision and recall** with an emphasis on the latter.

Future approach

COVID-19 MLIA · TASK 2 MULTILINGUAL SEMANTIC SEARCH

AN INFORMATION RETRIEVAL SYSTEM FOR COVID-19 DOCUMENTS IN SPANISH

José Alberto Mesa Murgado · Universidad de Jaén

Contact me

@murgado

jmurgado@ujaen.es

Universidad de Jaén