Overview of the "Multilingual Semantic Search" Task

Giorgio Maria Di Nunzio[1], Maria Eskevich[2]

[1]University of Padua, Italy
[2]CLARIN ERIC, The Netherlands

Covid-19 MLIA Task 2

Round 1 Virtual Meeting 12-14 January 2021

Overview

- Task description
- Collection and topics
- Participants and runs
- Pooling
- Relevance judgements: system and assessors
- Evaluation Metrics
- Systems scores
- Ongoing work

Task Description 1/2

The goal of the Multilingual Semantic Search task is to

“collect relevant information for the community, the **general public** as well as **other stakeholders**, when searching for health content in **different languages** and with **different levels of knowledge** about the specific topic.”

Task Description 2/2

There are two sub-tasks:

- Subtask 1 is a classic ad-hoc multilingual search task focused more on high precision.
  - Participant can submit a run with at most 1,000 documents retrieved per topic (for a total of 30,000 retrieved documents).

- Subtask 2 is more oriented towards high-recall systems, like Technology Assisted Review (TAR) systems.
  - Each run must have at most a total of 6,000 documents retrieved overall. In this second subtask, we expect on average 200 document retrieved per topic; however, participants can decide to distribute documents unevenly.

Participants can submit at most 5 monolingual runs (per language) and 5 bilingual runs (per pair of languages) per subtask.

Collections 1/2

In this first round, eight languages, seven of which are official EU languages, were chosen to build the initial set of collections of documents.

English, French, German, Greek, Italian, Spanish, Swedish, Ukranian

These languages are related to countries where COVID spread quickly or was managed in a different way at the beginning of 2020. For each language, five corpora where selected.

Collections 2/2

Language	Number of documents per corpora					TOTAL
	EU Press Corner	EUR-Lex	Global Voices	MEDISYS	Wikipedia	
English (en)	335	352	571	1 450 251	731	**1 452 240**
French (fr)	276	345	446	325 178	357	**326 599**
German (de)	266	345	51	272 645	364	**273 761**
Greek (el)	120	344	328	146 763	103	**147 658**
Italian (it)	123	342	539	661 514	271	**662 789**
Spanish (es)	115	342	595	832 639	342	**833 763**
Swedish (sv)	122	343	5	37 615	111	**38 196**
Ukrainian (uk)	0	0	66	15 395	121	**15 582**

Topics 1/2

The thirty topics used in Round 1 were created by selecting

1) a subset of the queries created for the TREC-COVID Task (courtesy of TREC-COVID Task organizers)

2) a selection of queries made available in the Bing search dataset for Coronavirus Intent which includes queries from all over the world issued on Bing that had an explicit/implicit intent related to the Coronavirus or Covid-19.

Topics 2/2

Topics are structured in the following way:

```
<topic number"topic identifier" xml:lang="ISO 639-1 code" >
    <keyword>surgical mask protection</keyword>
    <conversational>does a surgical mask protect from Covid-19?</conversational>
    <explanation>The documents retrieved should contain information about ...</explanation>
</topic>
```

Topics were translated from English to the other seven languages (plus Chinese and Japanese).

Participants 1/4

25 participants registered.

Four participants submitted runs for this task (listed in order of user id):

- Charles University, Czech Republic (**cunimtir**);
- University of Sheffield, UK (**gatenlp**);
- University of Padua, Italy (**ims**);
- Universidad de Jaén, Spain (**sinai**).

Participants 2/4

Subtask 1 - Number of monolingual / bilingual runs per participant

Language	cunimtir	gatenlp	ims	sinai	total
English	5 / 20	5 / 0	5 / 0	0 / 0	15 / 20
French	0 / 0	5 / 5	4 / 0	0 / 0	9 / 5
German	0 / 0	5 / 5	4 / 0	0 / 0	9 / 5
Greek	0 / 0	0 / 0	3 / 0	0 / 0	3 / 0
Italian	0 / 0	0 / 0	4 / 0	0 / 0	4 / 0
Spanish	0 / 0	5 / 5	4 / 0	5 / 0	14 / 0
Swedish	0 / 0	0 / 0	3 / 0	0 / 0	3 / 0
Ukrainian	0 / 0	0 / 0	3 / 0	0 / 0	3 / 0
total	5 / 20	20 / 15	30 / 0	5 / 0	60 / 35

Participants 3/4

Subtask 2 - Number of monolingual / bilingual runs per participant

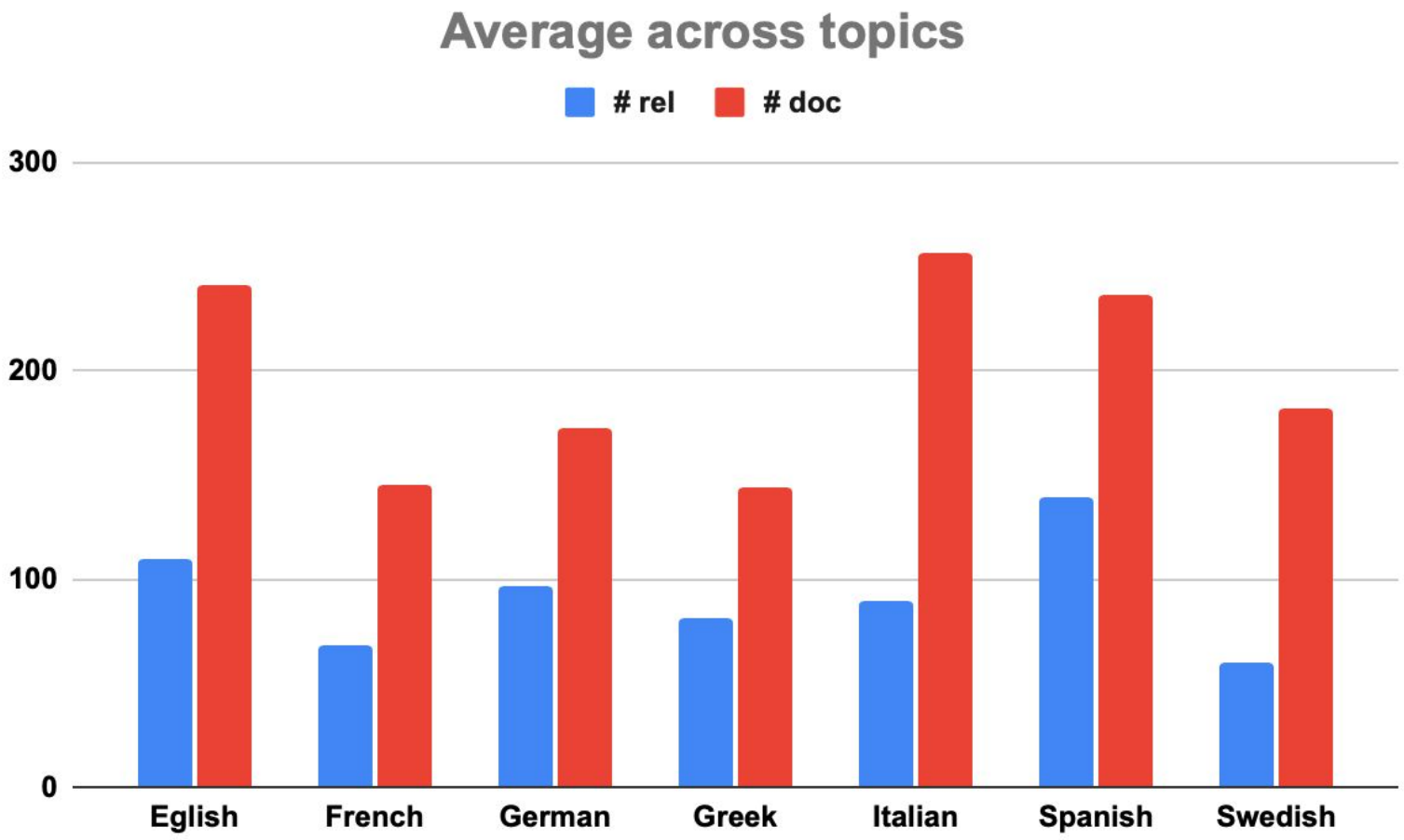
Language	cunimtir	gatenlp	ims	sinai	total
English	4 / 16	5 / 0	4 / 0	0 / 0	13 / 16
French	0 / 0	5 / 5	4 / 0	0 / 0	9 / 5
German	0 / 0	5 / 5	4 / 0	0 / 0	9 / 5
Greek	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
Italian	0 / 0	0 / 0	4 / 0	0 / 0	4 / 0
Spanish	0 / 0	5 / 5	4 / 0	5 / 0	14 / 5
Swedish	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
Ukrainian	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
total	4 / 16	20 / 15	20 / 0	5 / 0	49 / 31

Pooling

In order to build the pool of documents to be judged, we selected the **top k** documents for each run in order to reach, for each language, a pool of 6,000 - 8,000 unique documents.

Language	mono sub 1	mono sub 2	bili sub 1	bili sub 2	total	achieved	# of assessors
English	15	15	5	5	8,247	7,242	8
French	45	45	10	10	7,169	4,360	2
German	45	45	10	10	6,913	5,183	1
Greek	100	-	-	-	4,908	4,324	10
Italian	75	75	-	-	8,720	7,680	7
Spanish	23	23	5	5	7,091	7,091	4
Swedish	100	-	-	-	5,445	5,445	4

Relevance Judgements



Evaluation Metrics

- Subtask 1
  - Precision at 5
  - Normalized Discounted Cumulative Gain

- Subtask 2
  - Precision@k
  - Recall
  - RPrec

Systems main features

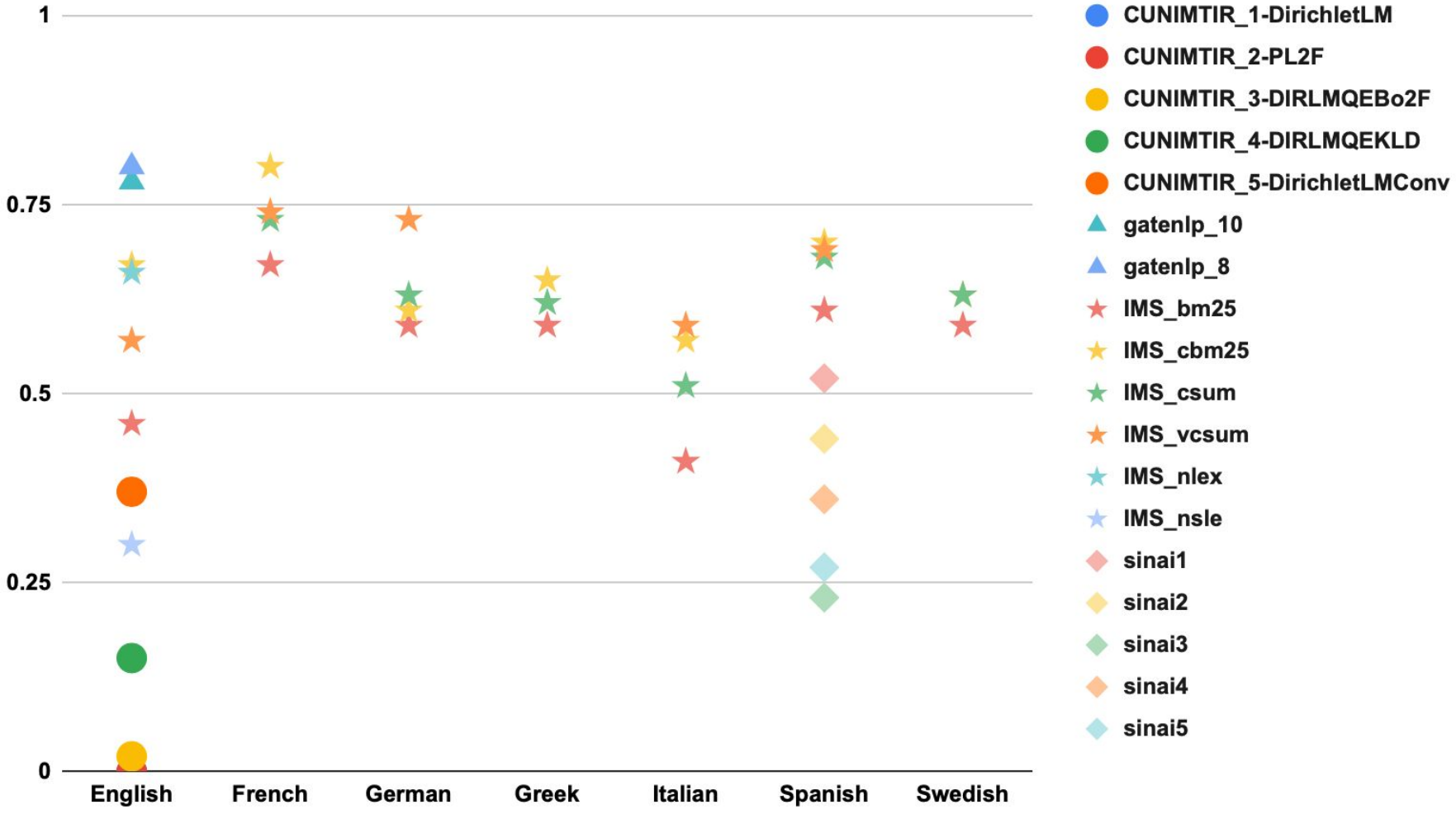
Core system:

- ElasticSearch
- Lucene
- Terrier

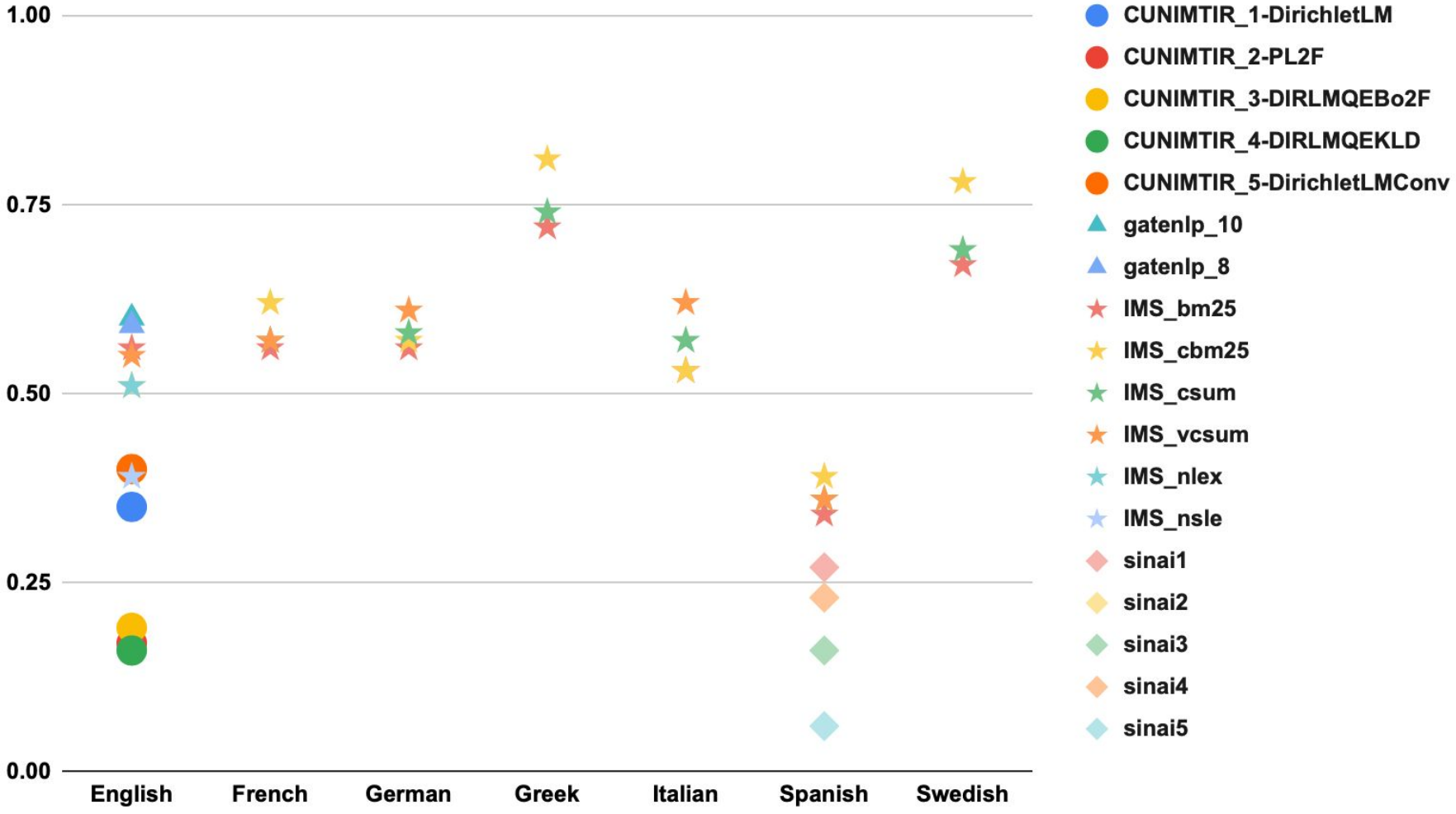
Approaches

- Keyword/Lexical Search
- Query Expansion/Reformulation
- Rank fusion
- Re-ranking neural models

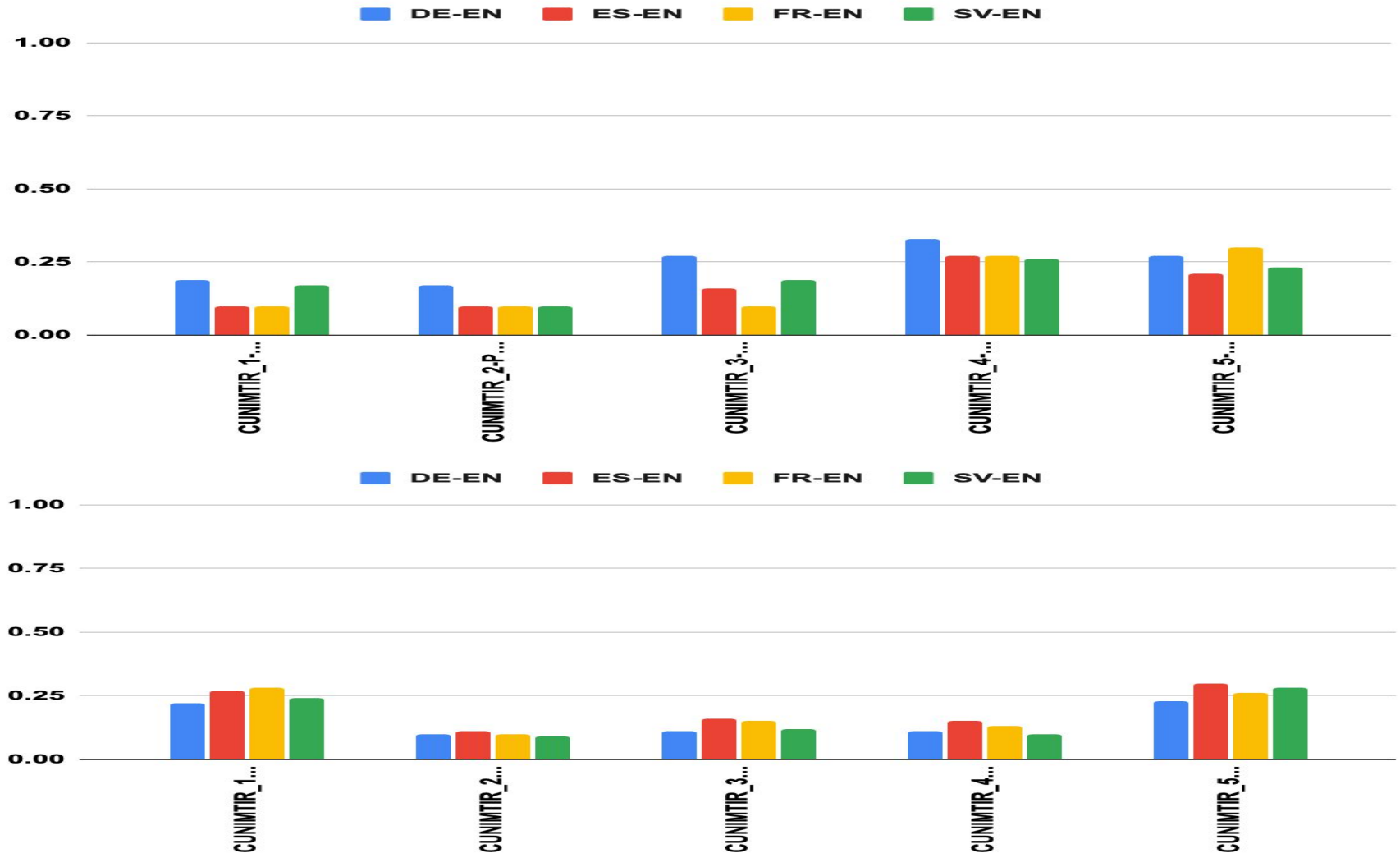
System Scores: Subtask 1
P5 for monolingual runs



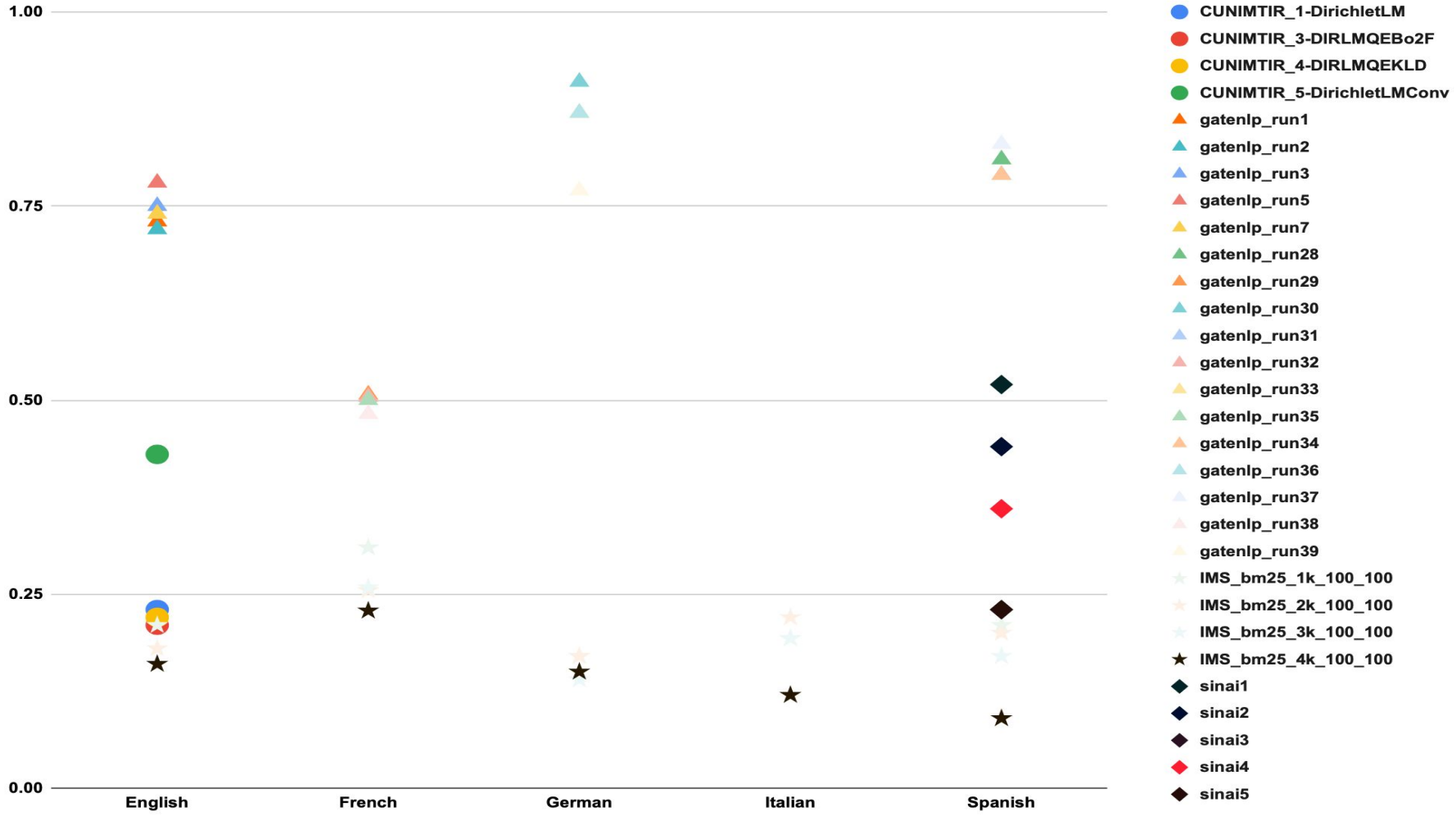
System Scores: Subtask 1
NDCG for monolingual runs



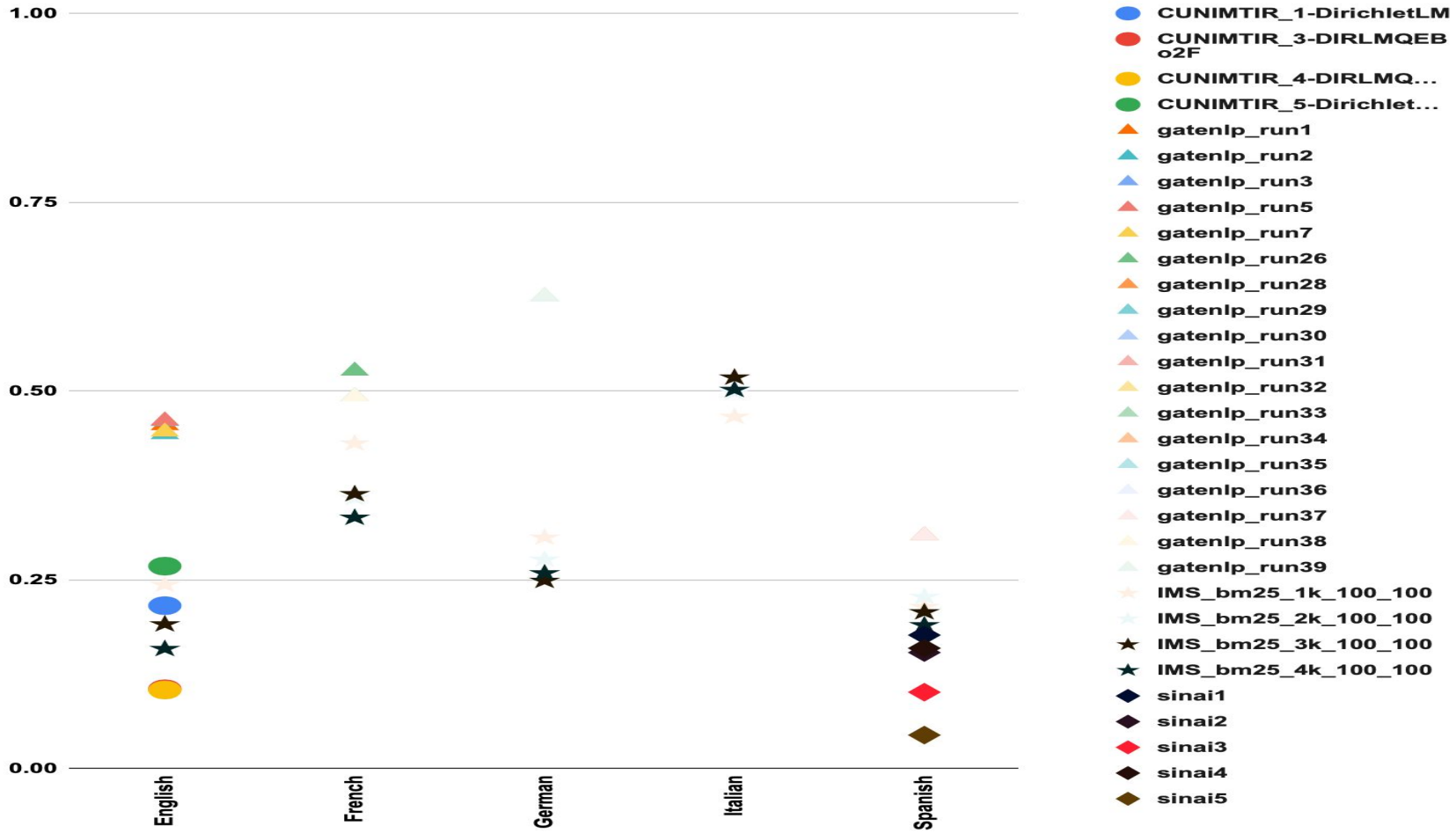
System Scores: Subtask 1
P5 and NDCG for bilingual runs



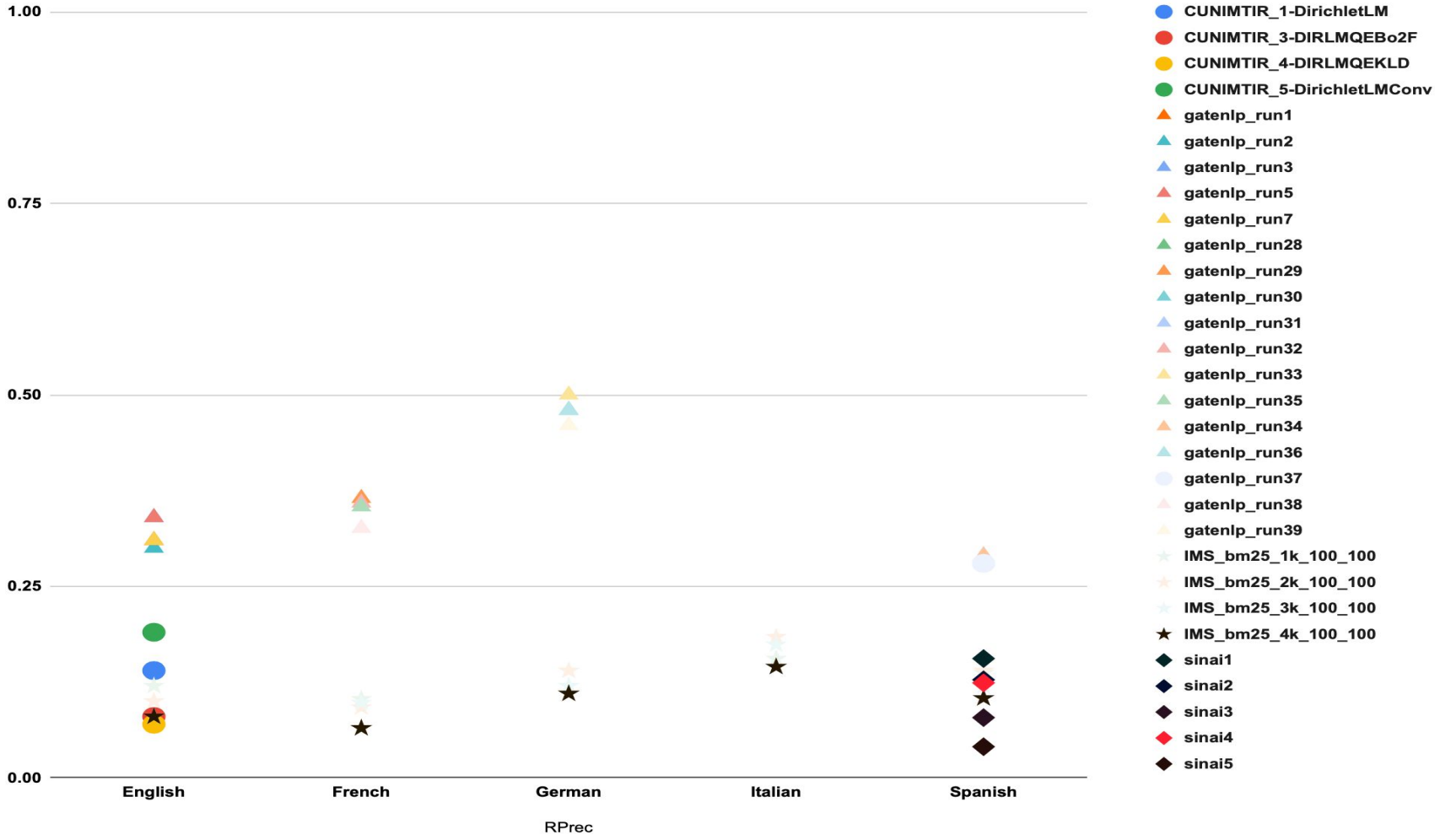
System Scores: Subtask 2
P5 for monolingual runs



System Scores: Subtask 2
Recall for monolingual runs



System Scores: Subtask 2
RPrec for monolingual runs



Ongoing work

Organisers:

- Assessment of runs for Ukranian
- Analysis of the results outliers that we have at this point, e.g. Swedish vs others
- Analysis of results per groups of topics, e.g. taken from different sets, type of information request

Together with the teams:

- Analysis of difference in performance based on the approach taken