

Analiza finalnih odgovora

Merlin Andrija, Lisica Ivan

2024-01-23

Radimo analizu finalnih odgovora na 30 pitanja. Iz 29 konfiguracija, svaki od 30 odgovora sadrži sljedeće podatke koji su bitni za analizu: - time - blue score - rouge score: precision, recall, f1 - human evaluation: expression and logic, accuracy and relevance, overall quality and engagement

Učitamo podatke

```
#ucitvamo sve podatke u listu podatkovnih okvira
file_list <- list.files(path = ".", pattern = "*.json", full.names = TRUE)
config_list <- list()
file_list

## [1] "./config_1.json"      "./config_10.json"     "./config_11.json"
## [4] "./config_12.json"     "./config_13.json"     "./config_14.json"
## [7] "./config_15.json"     "./config_16.json"     "./config_17.json"
## [10] "./config_18.json"     "./config_19.json"     "./config_2.json"
## [13] "./config_20.json"     "./config_21.json"     "./config_22.json"
## [16] "./config_23.json"     "./config_24.json"     "./config_25.json"
## [19] "./config_26.json"     "./config_27.json"     "./config_3.json"
## [22] "./config_4.json"      "./config_5.json"      "./config_6.json"
## [25] "./config_7.json"      "./config_8.json"      "./config_9.json"
## [28] "./config_parent1.json" "./config_parent2.json"

for (i in 1:length(file_list)) {
  file <- fromJSON(file_list[i])$answers
  file <- data.frame(file$question_number, file$time, file$evaluation$bleu_score, file$evaluation$rouge,
    config_list[[i]] <- file
}
config_list[[1]]$expression_and_logic <- factor(config_list[[1]]$expression_and_logic,
  levels = c("1", "1.5", "2", "2.5", "3", "3.5", "4", "4.5", "5"),
  ordered = TRUE)
config_list[[1]]$accuracy_and_relevance <- factor(config_list[[1]]$accuracy_and_relevance,
  levels = c("1", "1.5", "2", "2.5", "3", "3.5", "4", "4.5", "5"),
  ordered = TRUE)
config_list[[1]]$overall_quality_and_engagement <- factor(config_list[[1]]$overall_quality_and_engagement,
  levels = c("1", "1.5", "2", "2.5", "3", "3.5", "4", "4.5", "5"),
  ordered = TRUE)
```

Ispisujemo srednju vrijednost svih kategorija:

```
#ispisujemo srednju vrijednost svih kategorija
ocekivanje <- lapply(config_list, function(x) {
  sapply(x, function(y) {
    mean(as.numeric(y))
  })
})
```

[illegible]

```
## Warning in mean(as.numeric(y)): NAs introduced by coercion
```

```
## Warning in mean(as.numeric(y)): NAs introduced by coercion
```

```
ocekivanje <- as.data.frame(ocekivanje)
ocekivanje <- ocekivanje[-1,]
colnames(ocekivanje) <- c("config_1","config_10","config_11","config_12","config_13",
    "config_14","config_15","config_16","config_17","config_18","config_19",
    "config_2","config_20","config_21","config_22","config_23","config_24",
    "config_25","config_26","config_27","config_3",
    "config_4","config_5","config_6","config_7","config_8","config_9", "config_pa
    "config_parent2")
```

```
ocekivanje
```

##	config_1	config_10	config_11	config_12
## file.time	66.5450000	84.1126667	132.0510000	39.1340000
## file.evaluation.bleu_score	0.1405987	0.1536526	0.1432737	0.09702275
## precision	0.2719785	0.2771072	0.2826628	0.28091803
## recall	0.3070516	0.3166548	0.3029405	0.25172807
## f1	0.2607190	0.2551571	0.2586946	0.21283358
## file.evaluation.diversity	0.9770378	0.9451463	0.9450705	0.91675025
## expression_and_logic	8.5333333	4.6333333	4.5333333	4.5000000
## accuracy_and_relevance	7.2000000	4.2000000	4.1666667	3.6666667
## overall_quality_and_engagement	7.2666667	4.1000000	3.8666667	3.6333333
##	config_13	config_14	config_15	config_16
## file.time	63.8176667	103.2686667	30.7163333	89.4170000
## file.evaluation.bleu_score	0.1173203	0.1443221	0.1410021	0.1781632
## precision	0.2732153	0.2913368	0.2760135	0.2890328
## recall	0.2761418	0.3319495	0.2626451	0.3230017
## f1	0.2360943	0.2787334	0.2294765	0.2698516
## file.evaluation.diversity	0.9165228	0.9710667	0.9217279	0.9445571
## expression_and_logic	4.7000000	4.5666667	4.6666667	4.5333333
## accuracy_and_relevance	4.2333333	4.1333333	3.8000000	3.8666667
## overall_quality_and_engagement	4.0333333	4.0666667	3.7666667	3.7666667
##	config_17	config_18	config_19	config_2
## file.time	235.4840000	51.2470000	65.0566667	192.3570000
## file.evaluation.bleu_score	0.1684152	0.1149219	0.1241085	0.1333278
## precision	0.3000582	0.2718421	0.2644181	0.2841638
## recall	0.3167277	0.2599637	0.2754378	0.3032107
## f1	0.2650807	0.2329508	0.2374296	0.2533930
## file.evaluation.diversity	0.9762983	0.9255431	0.9789061	0.9609918
## expression_and_logic	4.3333333	4.7000000	4.6333333	4.5000000
## accuracy_and_relevance	3.8000000	4.0000000	3.8000000	3.5666667
## overall_quality_and_engagement	3.5333333	3.9000000	3.6666667	3.5666667
##	config_20	config_21	config_22	config_23
## file.time	135.9646667	44.9683333	51.6206667	101.8110000
## file.evaluation.bleu_score	0.1388235	0.08932692	0.1081324	0.1278615
## precision	0.2876330	0.26179593	0.2599065	0.2929888
## recall	0.3037571	0.25217184	0.2729479	0.3009453
## f1	0.2552418	0.21463106	0.2266388	0.2637181
## file.evaluation.diversity	0.9453378	0.91773910	0.9185782	0.9861647
## expression_and_logic	4.4000000	4.5000000	4.6666667	4.4333333
## accuracy_and_relevance	3.9333333	3.7000000	3.8666667	3.8333333
## overall_quality_and_engagement	3.5333333	3.5000000	3.8000000	3.7333333

##	config_24	config_25	config_26	config_27
## file.time	30.71866667	97.8340000	204.3756667	50.25766667
## file.evaluation.bleu_score	0.08178288	0.1368849	0.1435743	0.08550878
## precision	0.25323358	0.2755484	0.2907648	0.25967801
## recall	0.24609873	0.3048561	0.2989257	0.26200384
## f1	0.21200125	0.2520362	0.2559777	0.21645963
## file.evaluation.diversity	0.92705780	0.9436299	0.9790664	0.92034195
## expression_and_logic	4.50000000	4.6000000	4.6333333	4.56666667
## accuracy_and_relevance	3.76666667	3.9666667	3.8333333	3.76666667
## overall_quality_and_engagement	3.60000000	3.7333333	3.6666667	3.66666667
##	config_3	config_4	config_5	config_6
## file.time	39.7300000	53.5213333	98.3006667	29.95766667
## file.evaluation.bleu_score	0.0927931	0.1082026	0.1376816	0.07427984
## precision	0.2777923	0.2661008	0.2797470	0.27373041
## recall	0.2670281	0.2744313	0.2883045	0.24752817
## f1	0.2155899	0.2337492	0.2479850	0.20761083
## file.evaluation.diversity	0.9132980	0.9177890	0.9545233	0.95684723
## expression_and_logic	4.6000000	4.6333333	4.6666667	4.63333333
## accuracy_and_relevance	3.7333333	4.0000000	4.0666667	3.63333333
## overall_quality_and_engagement	3.6333333	3.8333333	3.9333333	3.50000000
##	config_7	config_8	config_9	
## file.time	90.7496667	187.7790000	63.89900000	
## file.evaluation.bleu_score	0.1534384	0.1527259	0.09703248	
## precision	0.2681331	0.2984042	0.27817471	
## recall	0.2902554	0.3157777	0.25693212	
## f1	0.2453083	0.2684531	0.21623200	
## file.evaluation.diversity	0.9466677	0.9772701	0.91683640	
## expression_and_logic	4.7000000	4.6000000	4.70000000	
## accuracy_and_relevance	4.1000000	3.7666667	3.76666667	
## overall_quality_and_engagement	3.9000000	3.6666667	3.66666667	
##	config_parent1	config_parent2		
## file.time	80.2300000	94.7403333		
## file.evaluation.bleu_score	0.1169219	0.1285883		
## precision	0.3228032	0.2749796		
## recall	0.2976971	0.2858366		
## f1	0.2661522	0.2410554		
## file.evaluation.diversity	0.9419697	0.9062623		
## expression_and_logic	4.6666667	4.6000000		
## accuracy_and_relevance	4.0000000	3.9000000		
## overall_quality_and_engagement	3.8666667	3.7333333		