

TREC 2018 Precision Medicine - Medical University of Graz

Pablo López-García

pablo.lopez@medunigraz.at

Institute for Medical Informatics, Statistics and Documentation

Medical University of Graz

Auenbruggerplatz 2 - 8036 Graz (Austria)

ABSTRACT

In this paper we report on our participation in the TREC 2018 Precision Medicine track (team name: *imi_mug*). We submitted 5 fully automatic runs to both the *biomedical articles* and *clinical trials* subtasks. Our system was based on Elasticsearch, templates, and parameter grid search query generation, building heavily on our previous participation and the reference standard from 2017. Our results were well above the median for the biomedical articles subtask and median/below median for the clinical trials subtask.

INTRODUCTION

The goal of the TREC 2018 Precision Medicine track was to improve search for clinicians treating cancer patients. 50 cases describing potential patients, termed *topics*, were provided as input and contained three features or dimensions: *disease*, *gene*, and *demographic*. Very similarly to last year, the challenge was divided into two subtasks that consisted of retrieving (1) relevant *biomedical articles* for treatment (from PubMed as well as ASCO and AACR conference proceedings) and (2) relevant *clinical trials* for enrollment (from ClinicalTrials.gov). The description of the challenge, the datasets, the relevance judgment guidelines are available online¹.

¹<http://www.trec-cds.org/2018.html>

SYSTEM OVERVIEW & GENERAL APPROACH

This year we built heavily on our participation from last year. In particular, (1) we reused our indexing infrastructure, (2) we took advantage of the official reference standard released after the conference, and (3) we applied our learnt lessons. A description of our previous participation² and the overview paper from 2017³ are useful to better understand this paper.

We first reindexed all documents in Elasticsearch 5.4.0 and fixed some mistakes we had discovered, like only having indexed the *Conclusion* field in structured abstracts from PubMed. We used Elasticsearch's default options for indexing, except for the *clinical trials*, where we added shingles (word n-grams) indexing (size 2 and 3)⁴. Due to the time and index size required for this task (2GB for 241K clinical trials alone) we decided not to index the biomedical articles using n-grams. **Appendix A** contains a summary of our final indices and some sample documents.

Our aim was to automatically generate Elasticsearch queries that produced good results using last year's datasets for reference and training.

²https://trec.nist.gov/pubs/trec26/papers/imi_mug-PM.pdf

³<https://trec.nist.gov/pubs/trec26/papers/Overview-PM.pdf>

⁴<https://www.elastic.co/guide/en/elasticsearch/reference/5.4/analysis-shingle-tokenfilter.html>

QUERY TEMPLATES, EXPERIMENTS & EVALUATION

To find those optimal queries, we reused our *query templates* from our previous participation. As a first step we analyzed our results from last year and reconsidered our strategies.

The main strategies we kept included: (1) exploring *must* and *should* clauses, (2) searching for *disease* and *gene* features in multiple fields and fine tuning *multi_match type*, *tie_breaker* and *boost* parameters, and (3) boosting specialized documents (extra abstracts from cancer conferences).

Strategies that we eliminated due to their poor results were: (1) positive and negative boosting of topic-oriented keywords, extra topics frequent keywords and chemotherapy suffixes, and (2) synonym expansion for diseases and genes.

New strategies introduced this year were: (1) mapping demographics in topics into MeSH age groups⁵, (2) splitting the list of genes into separate fields *gene1*, and *gene2*, (3) creating a specific index and mappings⁶ to be able to search using 2 and 3 word n-grams (for clinical trials only, due to the performance issues mentioned in the introduction), (4) evaluating the document's language, and (5) filtering documents with empty abstracts.

For convenience, we rewrote all our code in Python as Jupyter Notebooks⁷ (except the indexing part which we kept in Java), we used dataframes as data structures, and we relied on the *pytrec_eval* tool⁸ for evaluation, which we extended. This

⁵Custom variation of tree number M01.061 (0-2: newborn, 3-12: child, 13-18: adolescent, 19-44: adult, >44: aged).

⁶https://github.com/plopezgarcia/trec-2018-precision-medicine/blob/master/python/trials/trials_elastic_config.ipynb

⁷<http://jupyter.org/>

⁸https://github.com/cvangysel/pytrec_eval

restructuring of our framework was part of our future work and facilitated that we ran and evaluated experiments in an easier more replicable way (last year we used *Java*, *Travis*, and *trec_eval*).

Our algorithm to find those optimal queries via experiments was an iterative version of the following⁹:

1. Pick a query template.
2. Grid search the parameters.
3. Build and perform the associated Elasticsearch queries.
4. Calculate the aggregated evaluation measures for all the topics of 2017 using the official reference standard.
5. Check *recall at 1000*, *NDCG*, *R precision*, and *precision at 10* using the reference standard from 2017.
6. Save the template and optimized parameters as an Elasticsearch *json* query.
7. Perform the query for this year's topics.
8. Export results in standard *trec_eval* format as a run for submission.

SUBMITTED RUNS & RESULTS

We submitted 5 fully automatic runs to each subtask, where we tried to capture different strategies and evaluation measures. Our submissions, the exact steps we took, and the evaluation results for 2017 topics used for training are available online as Jupyter Notebooks for both the biomedical articles¹⁰ and the clinical trials¹¹ subtasks.

They can be summarized as follows:

⁹https://github.com/plopezgarcia/trec-2018-precision-medicine/blob/master/python/trec_utils/running.py

¹⁰https://github.com/plopezgarcia/trec-2018-precision-medicine/blob/master/python/abstracts/abstracts_submissions.ipynb

¹¹https://github.com/plopezgarcia/trec-2018-precision-medicine/blob/master/python/trials/trials_submission.ipynb

Biomedical Articles

1. Baseline well-performing query from 2017 with strict *multi_match (must)* of *disease* and *gene* in *title*, *abstract* and *MeSH tags* and grid-search optimized parameters *tie_breaker*, *type*, and *boost*. Relaxed matching / boosting (*should*) of demographics (*sex* and *age_group*) in MeSH tags: (relevance: 3).
2. Same as (1), plus boosting documents in English and belonging to the extra abstracts (relevance: 2).
3. Same as (2), plus search age group and sex in title, abstract, and MeSH (relevance: 4).
4. Same as (3), plus the abstract should exist (relevance: 5).
5. Same as (2), plus the abstract should exist (relevance: 1).

Clinical Trials

1. Baseline well-performing query from 2017 with strict *multi_match (must)* if *disease* and *gene* in *title*, *summary*, and *inclusion*, and relaxed matching/boosting in word n-grams fields. Relaxed matching/boosting of age range and sex (relevance: 3).
2. Same as (1) but the disease must not match the exclusion criteria (relevance: 5).
3. Same as (1), but all conditions are relaxed using *should* instead of *must* (relevance: 4).
4. Same as (1), but the relevance of the inclusion criteria weighs double (relevance: 2).
5. Same as (1), without grid search on word n-gram parameters for *gene1* and *gene2* (relevance: 1).

A comprehensive and more illustrative description of the results and how we obtained them can be found in the appendices.

Appendix B summarizes the submitted runs in a detailed table, and shows the exact Elasticsearch *json* queries that produced our submitted runs.

Appendix C shows our results as graphs and compares them with the average median and best results.

CONCLUSION

In this notebook we reported on our participation in the TREC 2018 Precision Medicine track and described our approach, strategies, results, and lessons learnt.

From our participation last year and the re-evaluation of our approaches using the reference standard, we learned that many promising strategies (like *disease* and *gene* expansion, and keyword boosting of chemotherapy suffixes) did not produce good results and we therefore removed them. Grid search was useful to fine-tune complex Elasticsearch query parameters, such as *multi_match type*, *tie_breaker* and *boost*.

Similarly to last year, we got substantially better results in the *biomedical articles* subtask, in good agreement with our effort spent in that task.

Rebuilding our Java, Travis, and *trec_eval*-based framework from last year as Python modules, using *pytrec_eval*, *dataframes*, and Jupyter notebooks proved to be time well spent and provided us with a solid environment where we can easily experiment in a completely replicable way.

APPENDIX A: Indices and Sample Documents

ALL BIOMEDICAL DATA - ELASTIC_SEARCH/_search?pretty

```
"hits":{
  "total":26980432,
}
```

PUBMED RECORDS - ELASTIC_SEARCH/abstracts/medline/_search?pretty

```
"hits":{
  "total":26669401,
  "max_score":1,
  "hits":[
    {
      "_index":"abstracts",
      "_type":"medline",
      "_id":"1832151",
      "_score":1,
      "_source":{
        "pubmedId":"1832151",
        "title":"Pseudomonas aeruginosa alkaline protease: evidence for secretion genes
and study of secretion mechanism.",
        "abstract":"A 6.5-kb DNA fragment carrying the functions required for specific
secretion of the extracellular alkaline protease produced by Pseudomonas aeruginosa was
cloned. The whole 6.5-kb DNA fragment was transcribed in one direction and probably carried
three genes involved in secretion. The expression in trans of these genes, together with the
apr gene, in Escherichia coli allowed synthesis and secretion of the alkaline protease,
which was extensively investigated by performing pulse-chase experiments under various
conditions. We demonstrated the absence of a precursor form, as well as the independence of
alkaline protease translocation from SecA. The absence of secretion genes impaired alkaline
protease secretion; the protein then remained intracellular and was partially degraded.",
        "publicationType":"Research Support, Non-U.S. Gov't",
        "language":"eng",
        "publicationYear":1991,
        "meshTags":[
          "Adenosine Triphosphatases",
```

```

    "Bacterial Proteins",
    "Blotting, Western",
    "Cloning, Molecular",
    "DNA, Bacterial",
    "Electrophoresis, Polyacrylamide Gel",
    "Escherichia coli",
    "Escherichia coli Proteins",
    "Genes, Bacterial",
    "Membrane Transport Proteins",
    "Plasmids",
    "Precipitin Tests",
    "Protein Precursors",
    "Pseudomonas aeruginosa",
    "Restriction Mapping",
    "SEC Translocation Channels",
    "Serine Endopeptidases"
  ],
  "medlineKeywords":[
  ]
}
},

```

EXTRA ABSTRACTS - `ELASTIC_SEARCH/abstracts/extra/_search?pretty`

```

"hits":{
  "total":70025,
  "max_score":1,
  "hits":[
    {
      "_index":"abstracts",
      "_type":"extra",
      "_id":"ASCO_72552-104",
      "_score":1,
      "_source":{
        "pubmedId":"ASCO_72552-104",
        "title":"Results of a phase II study of single-agent nab-paclitaxel in
platinum-refractory second-line metastatic urothelial carcinoma (UC).",
        "publicationDate":"2011",

```

```
"publicationYear":2011,
```

```
"abstract":"Background: There is currently no standard second-line chemotherapy for platinum- refractory UC. Both paclitaxel and docetaxel are commonly used but response rates are < 20% and no survival advantage has been shown. In this multi-institutional phase II study, we evaluated the efficacy and tolerability of a new albumin-bound nanoparticle formulation of paclitaxel, known as Abraxane (ABI-007) as a single agent in patients with platinum-refractory metastatic UC. Methods: Patients with measurable UC, progressing on or after first-line platinum-based chemotherapy were enrolled onto this two-stage trial. ABI-007 was given at 260 mg/m2 IV q3weekly until progression. Clinical evaluation, CBC and blood chemistries were performed every cycle with restaging CT scans every 2 cycles. Results: Accrual is now complete with 48 patients enrolled. Baseline characteristics: Male: Female 40:8; median age 68; ECOG Performance Status 0:1:2, 15:24:8. 248 cycles were delivered: median 5.5 cycles/pt with 17/48 pts (35%) requiring dose reductions. Most frequent adverse events (AE) were alopecia (12%), fatigue (12%), pain (12%), neuropathy (9%) and nausea (4%). The most frequent grade 3+ AE were pain (45%), hypertension (14%), fatigue (8%), joint stiffness (5%), neuropathy (4%) and weakness (4%). Forty patients are evaluable for response: 1 (2.5%) complete response (CR), 11 (28%) partial responses (PR), 9 (23%) stable disease (SD) and 20 (49%) progressive disease. One patient was inevaluable for response, 7 patients are too early for evaluation. Conclusions: Single-agent ABI-007 was well tolerated with a response rate (CR+PR) of 33% (12/36) and a clinical benefit rate (CR+PR+SD) of 58% (21/36), representing one of the highest reported response rates to date in the second-line UC setting. These results suggest further study of ABI-007 in urothelial carcinoma is warranted."
```

```
}
```

```
},
```

CLINICAL TRIALS - ELASTIC_SEARCH/trials/trials/_search?pretty

```
"hits":{
  "total":241006,
  "max_score":1,
  "hits":[
    {
      "_index":"trials",
      "_type":"trials",
      "_id":"NCT02661217",
      "_score":1,
      "_source":{
        "id":"NCT02661217",
        "title":"Comparison of Pre- and Post-discharge Initiation of LCZ696 Therapy in HFREF Patients After an Acute Decompensation Event",
        "summary":"To explore two modalities of treatment initiation (Predischarge, and Postdischarge) with LCZ696 in HFREF patients following stabilization after an ADHF episode.",
        "sex":[
          "female",
```

```

    "male"
  ],
  "minimum_age":18,
  "maximum_age":100,
  "inclusion":" 1. Patients hospitalized due to acute decompensated HF episode (ADHF) as primary diagnosis) and consistent Signs & Symptoms 2. Diagnosis of HF New York Heart Association class IItoIV and reduced ejection fraction: Left ventricular ejection fraction >= 40% at Screening 3. Patients did not receive any IV vasodilators (except nitrates), and/or any IV inotropic therapy from the time of presentation for ADHF to Randomization 4. Stabilized (while in the hospital) for at least 24 hours leading to Randomization. 5. Meeting one of the following criteria: Patients on any dose of ACEI or ARB at screening ACEI/ARB naïve patients and patients not on ACEI or ARB for at least 4 weeks before screening. ",
  "exclusion":" 1. History of hypersensitivity to the sacubitril, valsartan, or any ARBs, NEP inhibitors or to any of the LCZ696 excipients. 2. Symptomatic hypotension and/or a SBP below 110 mm Hg or SBP above 180 mm Hg prior to randomization 3. End stage renal disease at Screening; or estimated GFR below 30 mL/min/1.73 m2 (as measured by MDRD formula at Randomization. 4. Serum potassium above 5.4 mmol/L at Randomization. 5. Known history of hereditary or idiopathic angioedema or angioedema related to previous ACE inhibitor or ARB therapy 6. Severe hepatic impairment, biliary cirrhosis and cholestasis"
  }
},

```

APPENDIX B: Submitted Runs (Queries)

Biomedical Articles - Summary

		RUN				
FEATURE	QUERY	1	2	3	4	5
disease	fields	title abstract meshTags				
	match type	multi_match best_fields				
	bool	must				
	tie_breaker	0.4				
	boost	1.5				
gene	fields	title abstract meshTags				
	match type	multi_match cross_fields				
	bool	must				
	tie_breaker	0.4				
	boost	1				
sex	fields	MeSH tags		title abstract meshTags		MeSH tags
	match type	match		multi_match best_fields		match
	bool	should				
	tie_breaker	-		0.4		-
age	fields	meshTags (age group)		title abstract meshTags		meshTags (age group)
	match type	match		multi_match best_fields		match
	bool	should				
	tie_breaker	-		0.4		-
(document) language	fields	-	language			
	match type	-	match (English)			
	bool	-	should			
(document) type	fields	-	_type			
	match type	-	match (extra/conference abstracts)			
	bool	-	should			
Document abstract	bool	-			Filter: must exist	
FEATURE	QUERY	1	2	3	4	5
RUN						

Clinical Trials - Summary

		RUN				
FEATURE	QUERY	1	2	3	4	5
disease	fields	title summary inclusion			title summary inclusion^2	
	match type	multi_match best_fields				
	bool	must		should	must	
	tie_breaker	0.4				
	boost	1.3				
	fields	title.shingles (word n-grams) summary.shingles inclusion.shingles			title.shingles summary.shingles inclusion.shingles^2	title.shingles summary.shingles inclusion.shingles
	match type	multi_match				
	bool	should				
	fields	-	exclusion		-	
	match typ	-	match		-	
	bool	-	must_not		-	
gene	fields	title summary inclusion				
	match type	multi_match best_fields				
	bool	must		should	must	
	tie_breaker	0.4				
	boost	0.8				
gene1	fields	title.shingles (word n-grams) summary.shingles inclusion.shingles			title.shingles summary.shingles inclusion.shingles^2	title.shingles summary.shingles inclusion.shingles
	match type	multi_match best_fields				multi_match
	bool	should				
	tie_breaker	0.4				
	boost	1.5				
gene2	fields	title.shingles (word n-grams) summary.shingles inclusion.shingles			title.shingles summary.shingles inclusion.shingles^2	title.shingles summary.shingles inclusion.shingles
	match type	multi_match best_fields				multi_match
	bool	should				
	tie_breaker	0.3				
	boost	1				
sex	fields	sex				
	match type	match				
	bool	should				

	tie_breaker	-		0.4		-
age	fields	minimum_age				
	match type	lte (less than or equal)				
	bool	should				
	fields	maximum_age				
	match type	gte (greater than or equal)				
	bool	should				
FEATURE	QUERY	1	2	3	4	5
RUN						

Biomedical Articles - Queries for Submitted Runs

imi_mug_abs1.json

```
{
  "from":0,
  "size":1000,
  "query":{
    "bool":{
      "must":[
        {
          "multi_match":{
            "query":"{{disease}}",
            "fields":[
              "title",
              "abstract",
              "meshTags"
            ],
            "tie_breaker":0.4,
            "type":"best_fields",
            "boost":1.5
          }
        },
        {
          "multi_match":{
            "query":"{{gene}}",
            "fields":[
              "title",
              "abstract",
              "meshTags"
            ],
            "tie_breaker":0.4,
            "type":"cross_fields",
            "boost":1
          }
        }
      ],
      "should":[
        {
          "match":{
            "meshTags":"{{sex}}"
          }
        },
        {
          "match":{
            "meshTags":"{{age_group}}"
          }
        }
      ]
    }
  }
}
```

```
{
  "from":0,
  "size":1000,
  "query":{
    "bool":{
      "must":[
        {
          "multi_match":{
            "query":"{{disease}}",
            "fields":[
              "title",
              "abstract",
              "meshTags"
            ],
            "tie_breaker":0.4,
            "type":"best_fields",
            "boost":1.5
          }
        },
        {
          "multi_match":{
            "query":"{{gene}}",
            "fields":[
              "title",
              "abstract",
              "meshTags"
            ],
            "tie_breaker":0.4,
            "type":"cross_fields",
            "boost":1
          }
        }
      ],
      "should":[
        {
          "match":{
            "meshTags":"{{sex}}"
          }
        },
        {
          "match":{
            "meshTags":"{{age_group}}"
          }
        },
        {
          "match":{
            "language":"eng"
          }
        },
        {
          "match":{
            "_type":"extra"
          }
        }
      ]
    }
  }
}
```

```
{
  "from":0,
  "size":1000,
  "query":{
    "bool":{
      "must":[
        {
          "multi_match":{
            "query":"{{disease}}",
            "fields":[
              "title",
              "abstract",
              "meshTags"
            ],
            "tie_breaker":0.4,
            "type":"best_fields",
            "boost":1.5
          }
        },
        {
          "multi_match":{
            "query":"{{gene}}",
            "fields":[
              "title",
              "abstract",
              "meshTags"
            ],
            "tie_breaker":0.4,
            "type":"cross_fields",
            "boost":1
          }
        }
      ],
      "should":[
        {
          "multi_match":{
            "query":"{{sex}}",
            "fields":[
              "title",
              "abstract",
              "meshTags"
            ],
            "tie_breaker":0.4,
            "type":"best_fields"
          }
        },
        {
          "multi_match":{
            "query":"{{age_group}}",
            "fields":[
              "title",
              "abstract",
              "meshTags"
            ],
            "tie_breaker":0.4,
            "type":"best_fields"
          }
        }
      ],
      "match":{
        "language":"eng"
      }
    }
  }
}
```



```
        "abstract",
        "meshTags"
    ],
    "tie_breaker":0.4,
    "type":"cross_fields",
    "boost":1
}
},
"should":[
{
    "match":{
        "meshTags":"{{sex}}"
    }
},
{
    "match":{
        "meshTags":"{{age_group}}"
    }
},
{
    "match":{
        "language":"eng"
    }
},
{
    "match":{
        "_type":"extra"
    }
}
]
}
}
```


Clinical Trials - Queries for Submitted Runs

imi_mug_ct1.json

```
{
  "from":0,
  "size":1000,
  "query":{
    "bool":{
      "must":[
        {
          "multi_match":{
            "query":"{{disease}}",
            "fields":[
              "title",
              "summary",
              "inclusion"
            ],
            "tie_breaker":0.4,
            "type":"best_fields",
            "boost":1.3
          }
        },
        {
          "multi_match":{
            "query":"{{gene}}",
            "fields":[
              "title",
              "summary",
              "inclusion"
            ],
            "tie_breaker":0.4,
            "type":"best_fields",
            "boost":0.8
          }
        }
      ],
      "should":[
        {
          "range":{
            "minimum_age":{
              "lte":{
                {
                  age
                }
              }
            }
          }
        },
        {
          "range":{
            "maximum_age":{
              "gte":{
                {
                  age
                }
              }
            }
          }
        }
      ],
      "match":{
```



```

        "query":"{{disease}}",
        "fields":[
            "title",
            "summary",
            "inclusion"
        ],
        "tie_breaker":0.4,
        "type":"best_fields",
        "boost":1.3
    }
},
{
    "multi_match":{
        "query":"{{gene}}",
        "fields":[
            "title",
            "summary",
            "inclusion"
        ],
        "tie_breaker":0.4,
        "type":"best_fields",
        "boost":0.8
    }
},
],
"should":[
    {
        "range":{
            "minimum_age":{
                "lte":{
                    {
                        age
                    }
                }
            }
        }
    },
    {
        "range":{
            "maximum_age":{
                "gte":{
                    {
                        age
                    }
                }
            }
        }
    },
    {
        "match":{
            "sex":"{{sex}}"
        }
    }
],
"should":[
    {
        "multi_match":{
            "query":"{{disease}}",
            "fields":[
                "title.shingles",
                "summary.shingles",
                "inclusion.shingles"
            ]
        }
    }
],

```

```

    {
      "multi_match":{
        "query":"{{gene1}}",
        "fields":[
          "title.shingles",
          "summary.shingles",
          "inclusion.shingles"
        ],
        "tie_breaker":0.4,
        "type":"best_fields",
        "boost":1.5
      }
    },
    {
      "multi_match":{
        "query":"{{gene2}}",
        "fields":[
          "title.shingles",
          "summary.shingles",
          "inclusion.shingles"
        ],
        "tie_breaker":0.3,
        "type":"best_fields",
        "boost":1
      }
    }
  ]
}

```

imi_mug_ct3.json

```

{
  "from":0,
  "size":1000,
  "query":{
    "bool":{
      "should":[
        {
          "multi_match":{
            "query":"{{disease}}",
            "fields":[
              "title",
              "summary",
              "inclusion"
            ],
            "tie_breaker":0.4,
            "type":"best_fields",
            "boost":1.3
          }
        },
        {
          "multi_match":{
            "query":"{{gene}}",
            "fields":[
              "title",
              "summary",
              "inclusion"
            ],
            "tie_breaker":0.4,

```

```

        "type": "best_fields",
        "boost": 0.8
    }
},
],
"should": [
    {
        "range": {
            "minimum_age": {
                "lte": {
                    {
                        age
                    }
                }
            }
        }
    },
    {
        "range": {
            "maximum_age": {
                "gte": {
                    {
                        age
                    }
                }
            }
        }
    },
    {
        "match": {
            "sex": "{{sex}}"
        }
    }
],
"should": [
    {
        "multi_match": {
            "query": "{{disease}}",
            "fields": [
                "title.shingles",
                "summary.shingles",
                "inclusion.shingles"
            ]
        }
    },
    {
        "multi_match": {
            "query": "{{gene1}}",
            "fields": [
                "title.shingles",
                "summary.shingles",
                "inclusion.shingles"
            ],
            "tie_breaker": 0.4,
            "type": "best_fields",
            "boost": 1.5
        }
    },
    {
        "multi_match": {
            "query": "{{gene2}}",
            "fields": [
                "title.shingles",
                "summary.shingles",
                "inclusion.shingles"
            ]
        }
    }
]

```



```

"bool":{
  "must":[
    {
      "multi_match":{
        "query":"{{disease}}",
        "fields":[
          "title",
          "summary",
          "inclusion"
        ],
        "tie_breaker":0.4,
        "type":"best_fields",
        "boost":1.3
      }
    },
    {
      "multi_match":{
        "query":"{{gene}}",
        "fields":[
          "title",
          "summary",
          "inclusion"
        ],
        "tie_breaker":0.4,
        "type":"best_fields",
        "boost":0.8
      }
    }
  ],
  "should":[
    {
      "range":{
        "minimum_age":{
          "lte":{
            {
              age
            }
          }
        }
      }
    },
    {
      "range":{
        "maximum_age":{
          "gte":{
            {
              age
            }
          }
        }
      }
    },
    {
      "match":{
        "sex":"{{sex}}"
      }
    }
  ],
  "should":[
    {
      "multi_match":{
        "query":"{{disease}}",
        "fields":[
          "summary.shingles",
          "inclusion.shingles"
        ]
      }
    }
  ]
}

```



```
    ]
  },
  {
    "multi_match":{
      "query":"{{gene1}}",
      "fields":[
        "summary.shingles",
        "inclusion.shingles"
      ]
    }
  },
  {
    "multi_match":{
      "query":"{{gene2}}",
      "fields":[
        "summary.shingles",
        "inclusion.shingles"
      ]
    }
  }
]
}
```

APPENDIX C: Results

Biomedical Articles



Clinical Trials

