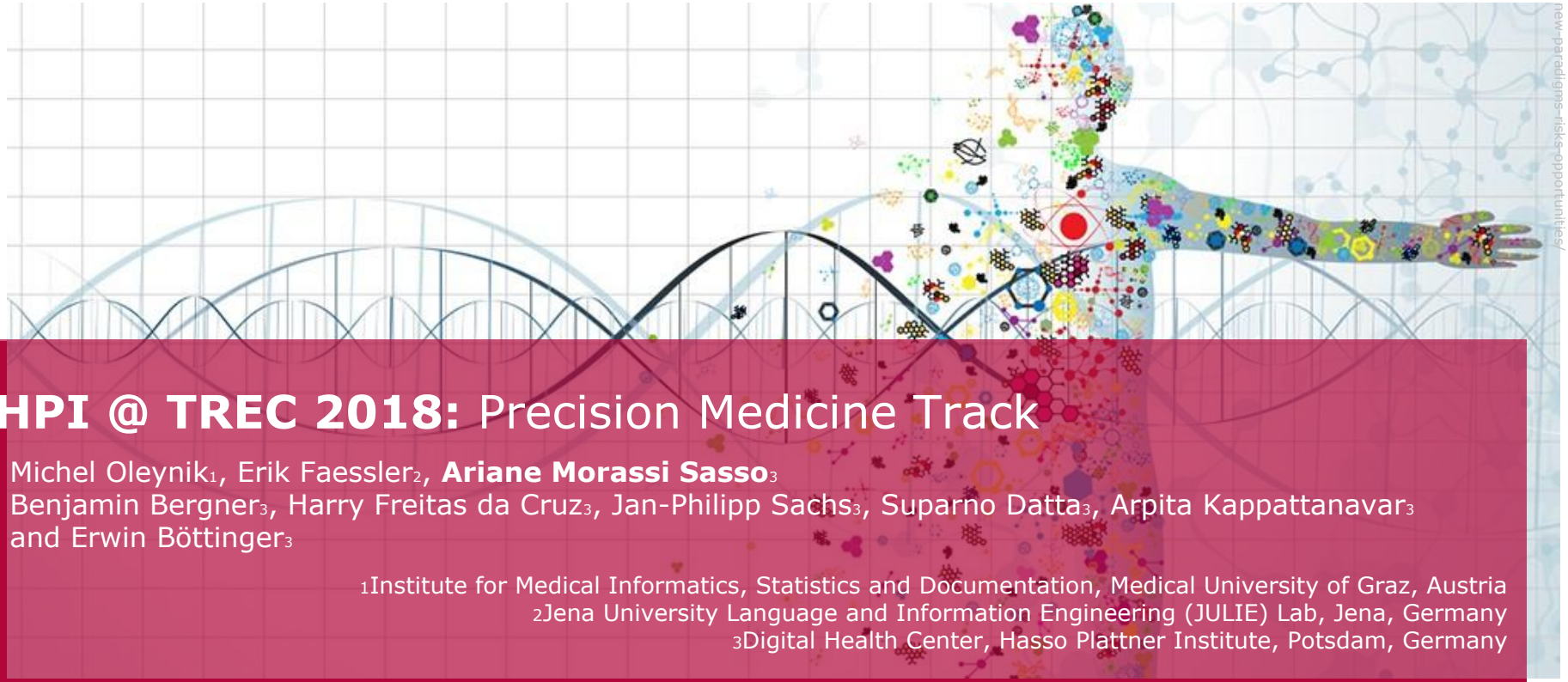


HPI @ TREC 2018: Precision Medicine Track

Michel Oleynik[1], Erik Faessler[2], **Ariane Morassi Sasso**[3]
Benjamin Bergner[3], Harry Freitas da Cruz[3], Jan-Philipp Sachs[3], Suparno Datta[3], Arpita Kappattanavar[3] and Erwin Böttinger[3]

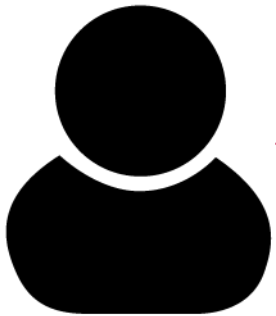
[1]Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria
[2]Jena University Language and Information Engineering (JULIE) Lab, Jena, Germany
[3]Digital Health Center, Hasso Plattner Institute, Potsdam, Germany

Agenda

1. **TREC: Precision Medicine Track**
2. **Framework**
   - Med Uni Graz Framework
3. **Improvements**
   - Weighting Hypernyms and Synonyms
   - TF-IDF and Topic Modeling (LDA) for Keyword Selection
   - Precision Medicine Classifier
4. **Results**
5. **Discussion and Future Work**

TREC: Precision Medicine Track

Patient Case (Topic)



- Melanoma
- BRAF (V600K)
- 64 years old & male

→ Biomedical Articles (PubMed/ASCO/ACCR)

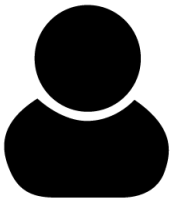
→ Clinical Trials (ClinicalTrials.gov)

TREC 2018

Ariane Sasso
16-11-2018

Med Uni Graz Framework

Query: Biomedical Articles



Melanoma

BRAF (V600K)

64 years old & male

+

Templates

Decorators

Vemurafenib in patients with BRAF(V600) mutated metastatic melanoma: an open-label, multicentre, safety study.

Larkin J[1], Del Vecchio M[2], Ascierto PA[3], Krajsova I[4], Schachter J[5], Neyns B[6], Espinosa E[7], Garbe C[8], Sileni VC[9], Gogas H[10], Miller WH Jr[11], Mandalà M[12], Hospers GA[13], Arance A[14], Queirolo P[15], Hauschild A[16], Brown MP[17], Mitchell L[18], Veronese L[18], Blank CU[19].

Author information

Abstract

BACKGROUND: The orally available BRAF kinase inhibitor vemurafenib, compared with dacarbazine, shows improved response rates, progression-free survival (PFS), and overall survival in patients with metastatic melanoma that has a BRAF(V600) mutation. We assessed vemurafenib in patients with advanced metastatic melanoma with BRAF(V600) mutations who had few treatment options.

METHODS: In an open-label, multicentre study, patients with untreated or previously treated melanoma and a BRAF(V600) mutation received oral vemurafenib 960 mg twice a day. The primary endpoint was safety. All analyses were done on the safety population, which included all patients who received at least one dose of vemurafenib. This report is the third interim analysis of this study. This study is registered with ClinicalTrials.gov, number NCT01307397.

Keywords

MeSH terms

Administration, Oral
Aged
Aged, 80 and over
Antineoplastic Agents/administration & dosage
Antineoplastic Agents/adverse effects
Antineoplastic Agents/therapeutic use*

TREC 2018

Ariane Sasso
16-11-2018

Chart **4**

Med Uni Graz Framework

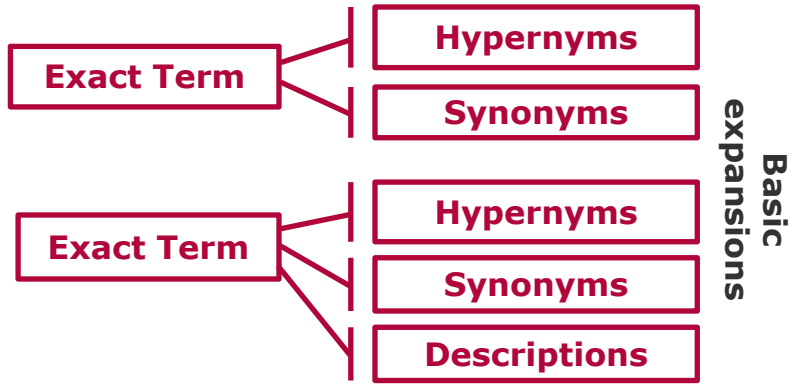
Query: Biomedical Articles

```
"bool": {
        "must": [
                {{biomedical_articles/disease.json}},
                {{biomedical_articles/gene.json}}
        ],
        "should": [
                {{biomedical_articles/extra.json}},
                {{biomedical_articles/chemotherapy.json}},
                {{biomedical_articles/cancer.json}},
                {{biomedical_articles/dna.json}},
                {{biomedical_articles/positive_boosters.json}},
                {{biomedical_articles/negative_boosters.json}},
                {{biomedical_articles/pm.json}}
        ],
        "must_not": [
                {{biomedical_articles/non_melanoma.json}}
        ]
}
```

Med Uni Graz Framework

Query: Biomedical Articles

```
"bool": {
        "must": [
                {{biomedical_articles/disease.json}},
                {{biomedical_articles/gene.json}}
        ],
        "should": [
                {{biomedical_articles/extra.json}},
                {{biomedical_articles/chemotherapy.json}},
                {{biomedical_articles/cancer.json}},
                {{biomedical_articles/dna.json}},
                {{biomedical_articles/positive_boosters.json}},
                {{biomedical_articles/negative_boosters.json}},
                {{biomedical_articles/pm.json}}
        ],
        "must_not": [
                {{biomedical_articles/non_melanoma.json}}
        ]
}
```



Med Uni Graz Framework

Query: Biomedical Articles

```
"bool": {
        "must": [
                {{biomedical_articles/disease.json}},
                {{biomedical_articles/gene.json}}
        ],
        "should": [
                {{biomedical_articles/extra.json}},
                {{biomedical_articles/chemotherapy.json}},
                {{biomedical_articles/cancer.json}},
                {{biomedical_articles/dna.json}},
                {{biomedical_articles/positive_boosters.json}},
                {{biomedical_articles/negative_boosters.json}},
                {{biomedical_articles/pm.json}}
        ],
        "must_not": [
                {{biomedical_articles/non_melanoma.json}}
        ]
}
```

Boosters

Med Uni Graz Framework

Query: Biomedical Articles

```
"bool": {
        "must": [
                {{biomedical_articles/disease.json}},
                {{biomedical_articles/gene.json}}
        ],
        "should": [
                {{biomedical_articles/extra.json}},
                {{biomedical_articles/chemotherapy.json}},
                {{biomedical_articles/cancer.json}},
                {{biomedical_articles/dna.json}},
                {{biomedical_articles/positive_boosters.json}},
                {{biomedical_articles/negative_boosters.json}},
                {{biomedical_articles/pm.json}}
        ],
        "must_not": [
                {{biomedical_articles/non_melanoma.json}}
        ]
}
```

Med Uni Graz Framework

Query: Biomedical Articles

```
"bool": {
        "must": [
                {{biomedical_articles/disease.json}},
                {{biomedical_articles/gene.json}}
        ],
        "should": [
                {{biomedical_articles/extra.json}},
                {{biomedical_articles/chemotherapy.json}},
                {{biomedical_articles/cancer.json}},
                {{biomedical_articles/dna.json}},
                {{biomedical_articles/positive_boosters.json}},
                {{biomedical_articles/negative_boosters.json}},
                {{biomedical_articles/pm.json}}
        ],
        "must_not": [
                {{biomedical_articles/non_melanoma.json}}
        ]
}
```

Med Uni Graz Framework

Query: Biomedical Articles

```
"bool": {
        "must": [
                {{biomedical_articles/disease.json}},
                {{biomedical_articles/gene.json}}
        ],
        "should": [
                {{biomedical_articles/extra.json}},
                {{biomedical_articles/chemotherapy.json}},
                {{biomedical_articles/cancer.json}},
                {{biomedical_articles/dna.json}},
                {{biomedical_articles/positive_boosters.json}},
                {{biomedical_articles/negative_boosters.json}},
                {{biomedical_articles/pm.json}}
        ],
        "must_not": [
                {{biomedical_articles/non_melanoma.json}}
        ]
}
```

PM Classifier

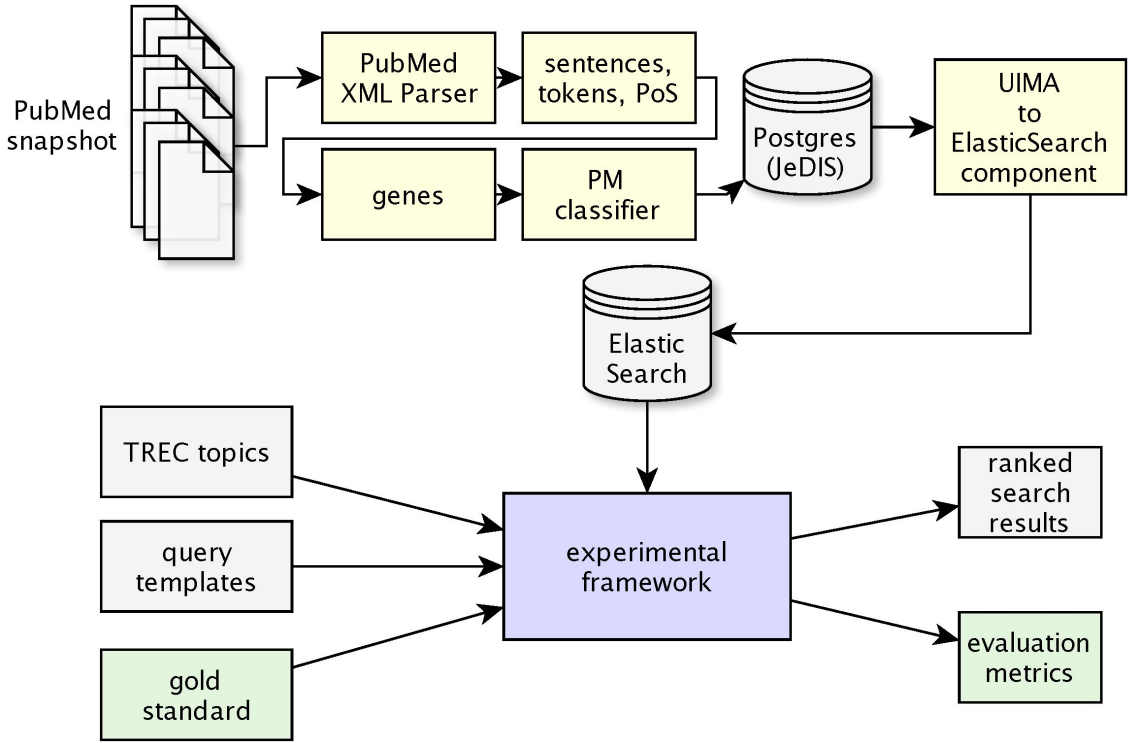
Med Uni Graz Framework

Query: Biomedical Articles

```
"bool": {
        "must": [
                {{biomedical_articles/disease.json}},
                {{biomedical_articles/gene.json}}
        ],
        "should": [
                {{biomedical_articles/extra.json}},
                {{biomedical_articles/chemotherapy.json}},
                {{biomedical_articles/cancer.json}},
                {{biomedical_articles/dna.json}},
                {{biomedical_articles/positive_boosters.json}},
                {{biomedical_articles/negative_boosters.json}},
                {{biomedical_articles/pm.json}}
        ],
        "must_not": [
                {{biomedical_articles/non_melanoma.json}}
        ]
}
```

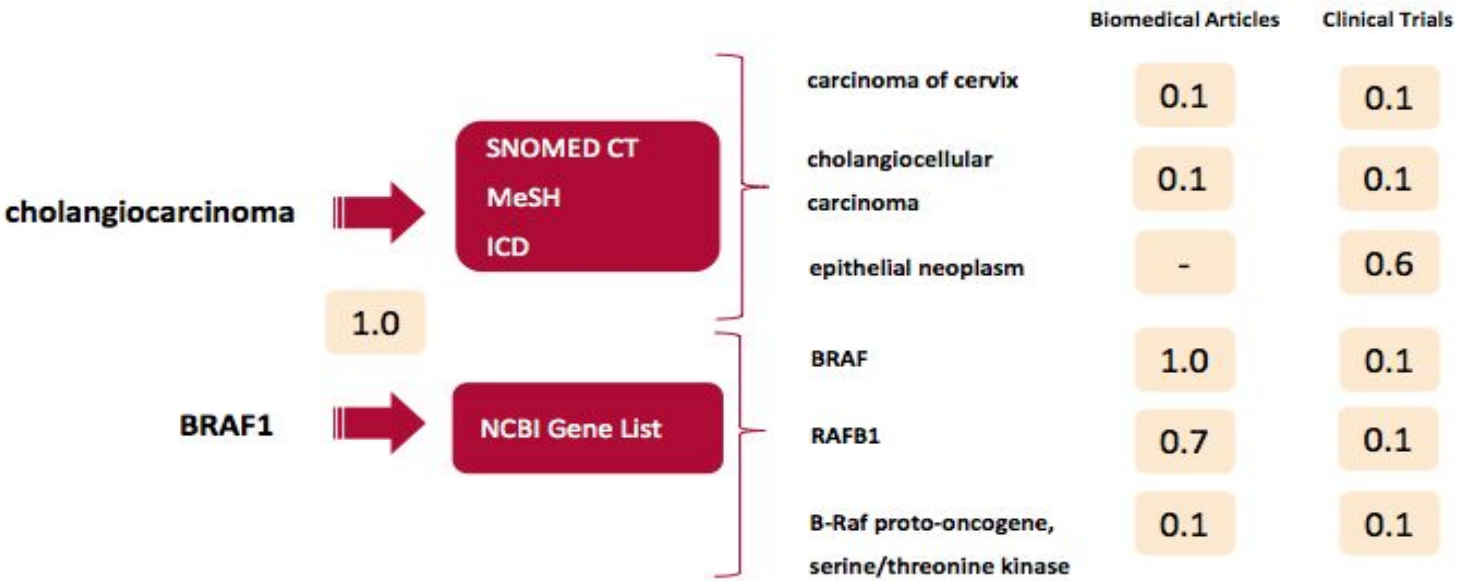
Hand-crafted Rules

Overview


PubMed snapshot
PubMed XML Parser
sentences, tokens, PoS
genes
PM classifier
Postgres (JeDIS)
UIMA to ElasticSearch component
Elastic Search
TREC topics
query templates
gold standard
experimental framework
ranked search results
evaluation metrics

Improvements

Weighting Hypernyms, Preferred Terms, Synonyms and Descriptions



Improvements

TF-IDF and **Topic Modeling (LDA)** for Keywords Selection

Gold Standard 2017

PM or NOT PM

Title
Abstracts
Mesh Terms

TF-IDF → **PM Terms**, **Not PM Terms**

LDA → **PM Topic** → **PM Terms**; **Not PM Topic** → **Not PM Terms**

TREC 2018

Ariane Sasso
16-11-2018

Improvements

TF-IDF and **Topic Modeling (LDA)** for Keywords Selection

Gold Standard 2017

PM or NOT PM

Title
Abstracts
Mesh Terms

TF-IDF

PM Terms

```
Topic 0
Word 0: patient
Word 1: cancer
Word 2: treatment
Word 3: study
Word 4: tumor
Word 5: year
Word 6: survival
Word 7: case
Word 8: disease
Word 9: risk
Word 10: therapy
```

LDA

PM Topic → PM Terms

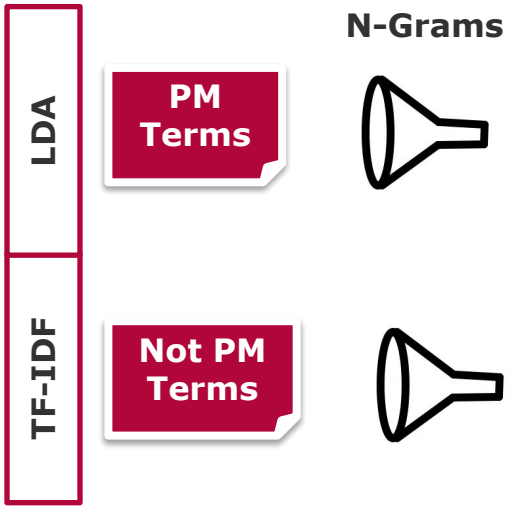
Not PM Topic → Not PM Terms

TREC 2018

Ariane Sasso
16-11-2018

Improvements

TF-IDF and **Topic Modeling (LDA)** for Keywords Selection



gefitinib treatment survival prognostic clinical prognosis therapy outcome resistance Gleason targets

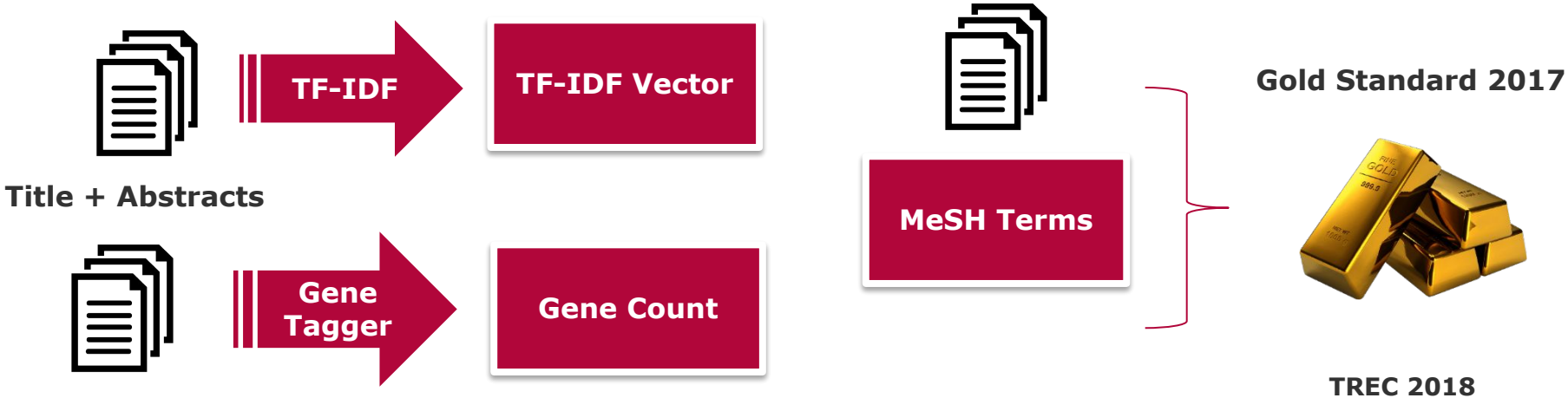
1

dna tumor cell mouse model tissue development specific staining pathogenesis case

-1

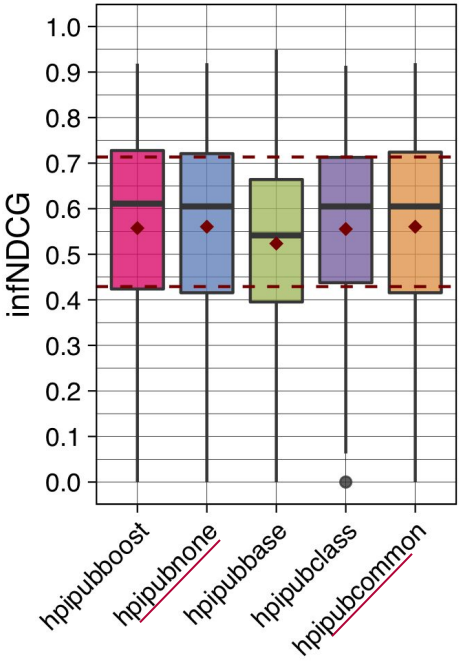
Improvements

PM Classifier

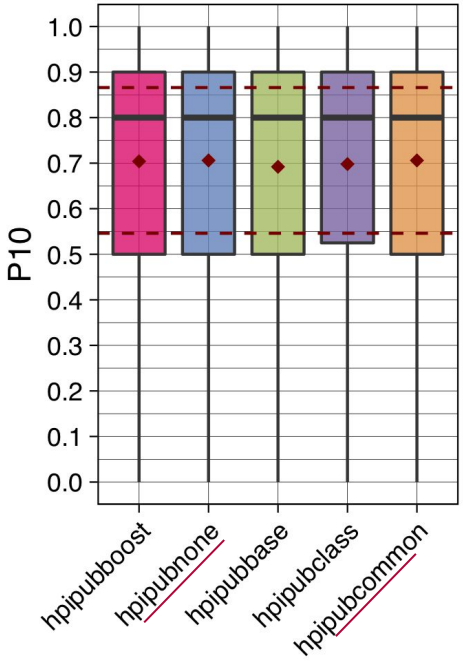


Accuracy: 75% (10-fold cross-validation)

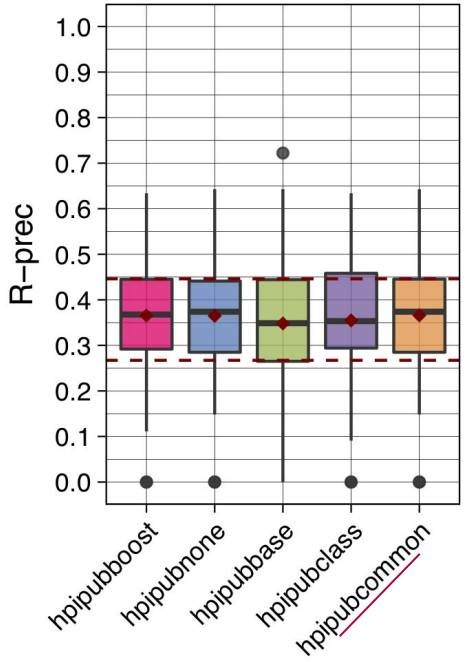
Results: Biomedical Articles



0.5605



0.7060



0.3658

Expansion

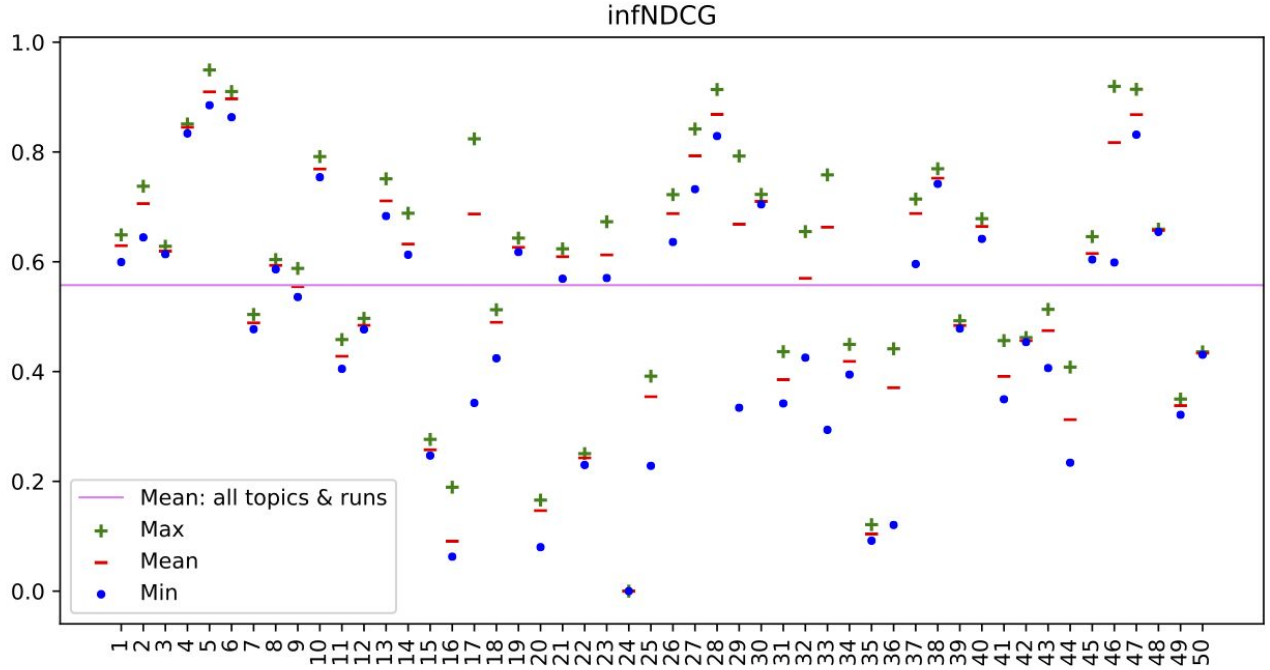
Classifier

TREC 2018

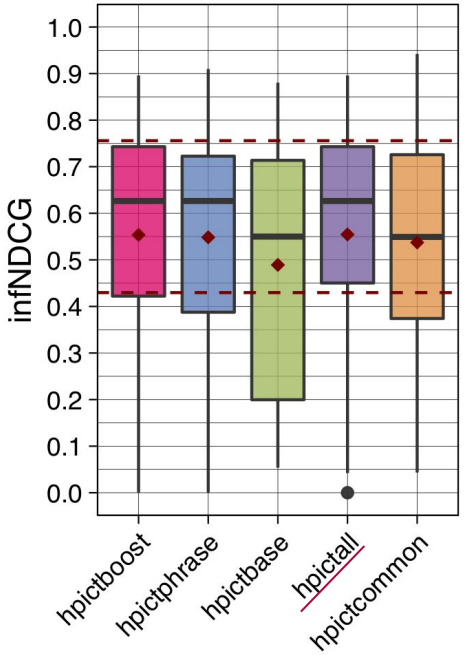
Ariane Sasso
16-11-2018

Chart **18**

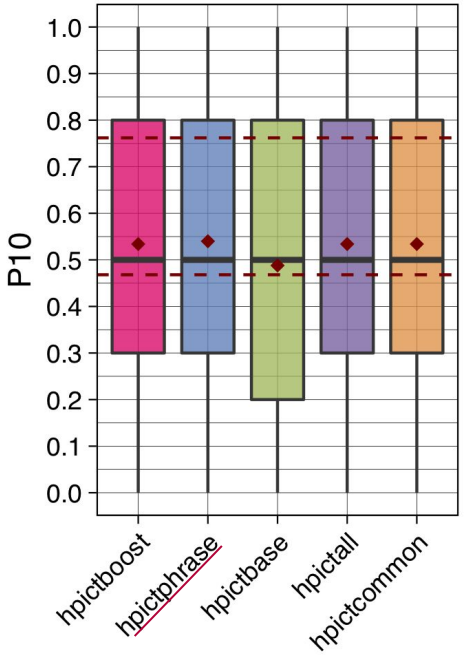
Biomedical Articles: infDNCG x Topic



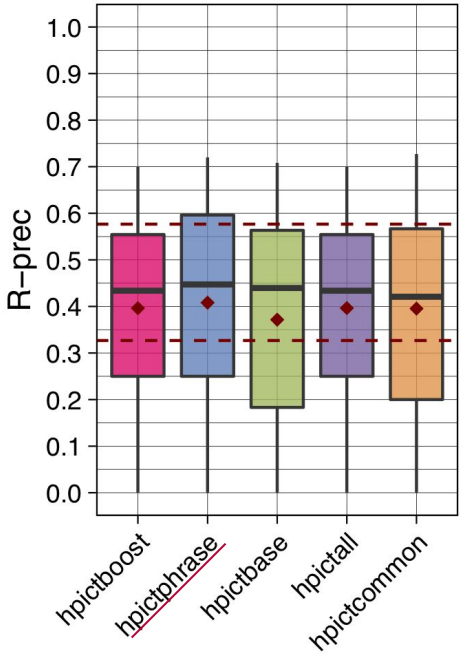
Results: Clinical Trials



0.5545



0.5400



0.4081

Solid Tumor Gene Family

TREC 2018

Ariane Sasso
16-11-2018

Discussion and Future Work

- **Supervised approach did not improve results**
  - Overlap with keyword selection
  - Did not test with Clinical Trials
- **Grid search for optimal weighting**
- **Better terminology coverage**
  - NCIt `Neoplasm_Has_Special_Category` maps solid tumors
- **Negation detection**
  - Topics 21/22: "no" x "extensive" "tumor infiltrating lymphocytes"

Team

Erik Faessler

Jena University, Germany

Language and Information Engineering (JULIE) Lab

Michel Oleynik

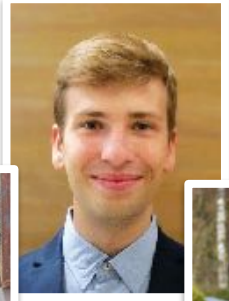
Medical University of Graz, Austria

Institute for Medical Informatics, Statistics and Documentation

Arpita Kappattanavar

Ariane Sasso

Benjamin Bergner



Suparno Datta

Harry Freitas da Cruz

Jan-Philipp Sachs

Prof. Dr. Böttinger

TREC 2018

Ariane Sasso
16-11-2018

Thank you!

Github:
https://github.com/hpi-dhc/trec-pm

Ariane Sasso
PhD Student
Hasso Plattner Institute