

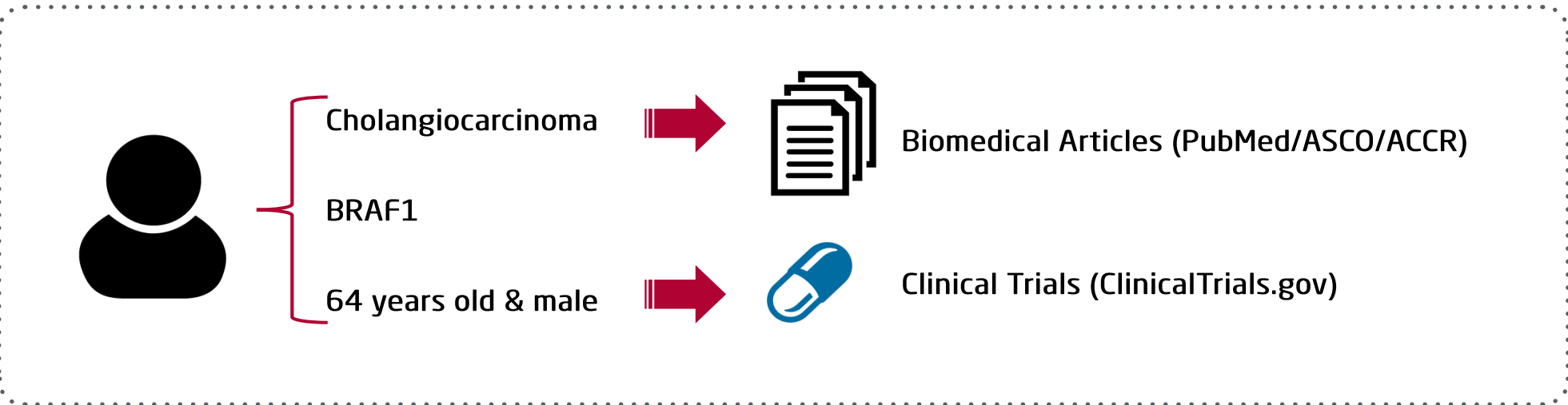
HPI @ TREC 2018

Precision Medicine Track

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

INTRODUCTION

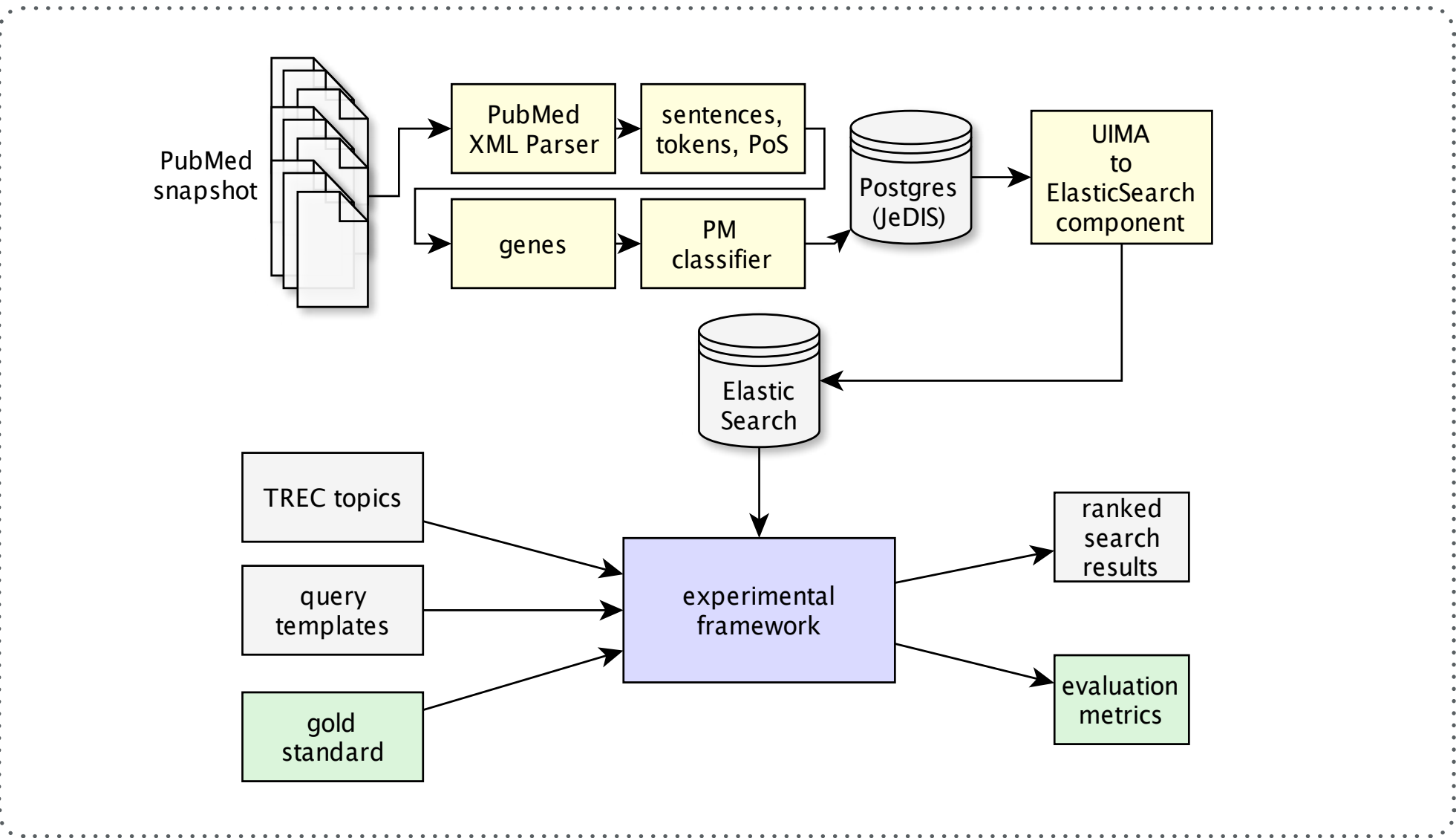
From 50 oncology cases (topics), retrieve relevant Biomedical Articles and Clinical Trials:



FRAMEWORK

Based on reusable components:

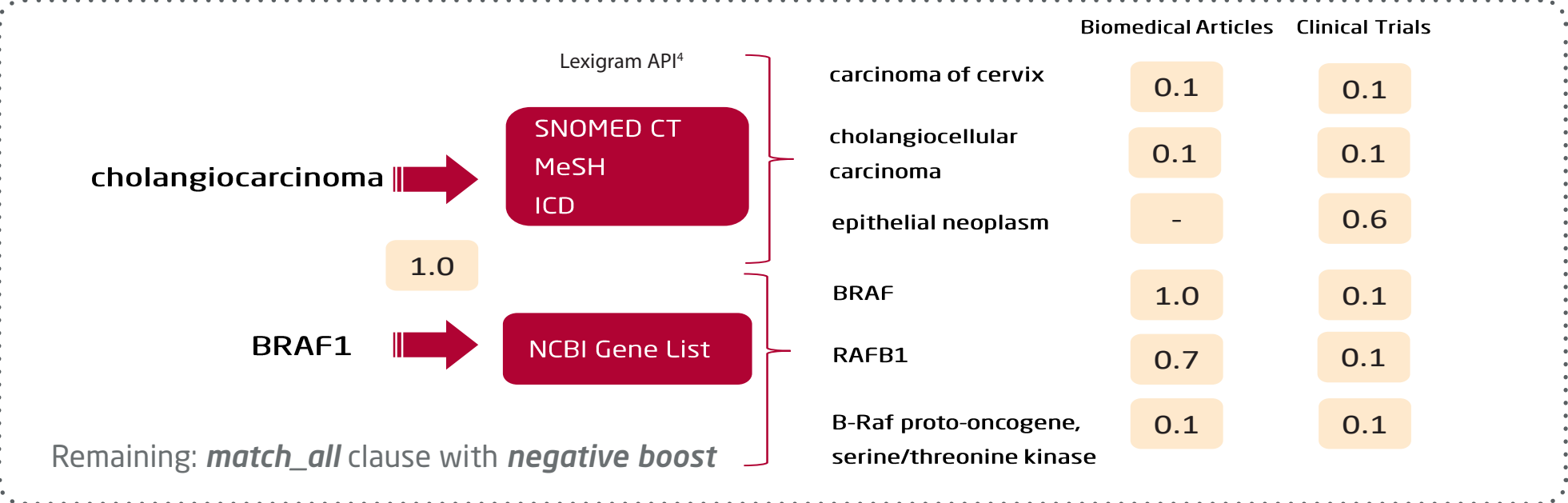
1. UIMA (Unstructured Information Management Architecture)¹
2. ElasticSearch (ES) 5.4.0 Server²
3. Experimental Java framework proposed by the Medical University of Graz (2017)³



STRATEGIES

Runs contained one or more of the following approaches:

1. Hypernyms and Synonyms Weighting (using disjunctive queries)
2. TF-IDF and Topic Modeling (LDA) for Keywords Selection
3. Precision Medicine Classifier trained on the TREC-PM 2017 Gold Standard
4. Hand-Crafted Rules (e.g.: solid tumors, non-melanoma)

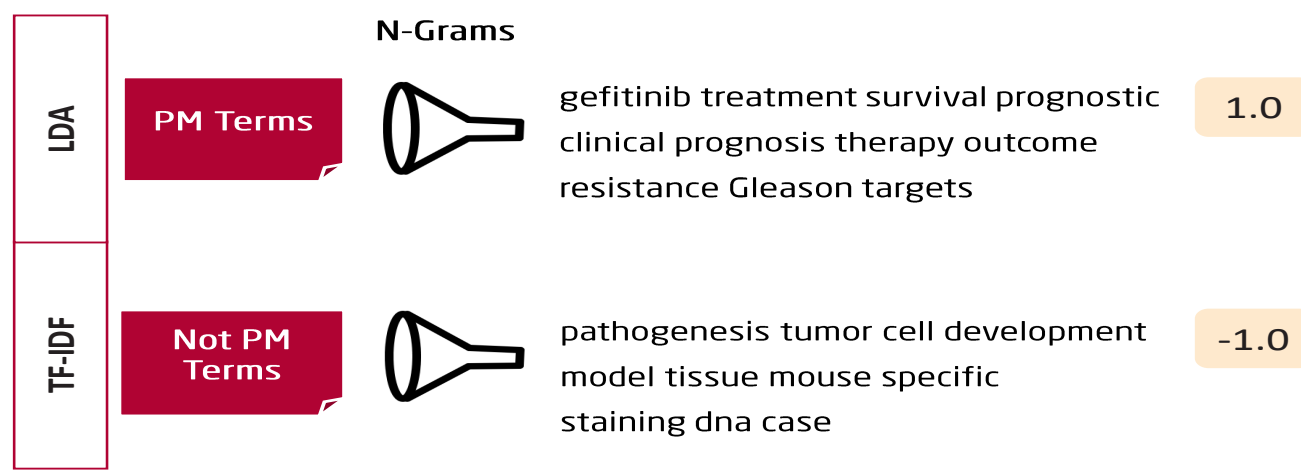


REFERENCES

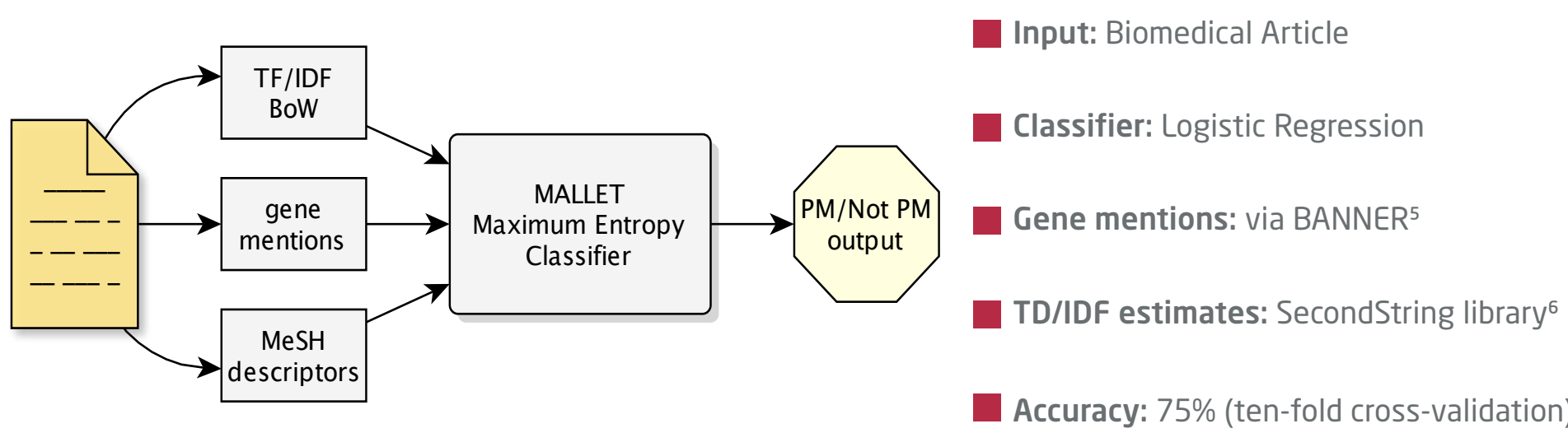
1. <https://uima.apache.org>
2. <https://www.elastic.co>
3. <https://github.com/bst-mug/trec-pm>
4. <https://www.lexigram.io>
5. <http://banner.sourceforge.net>
6. <http://www.secondstring.sourceforge.net>

■ Topic Modeling

Training: TREC-PM 2017 Gold Standard

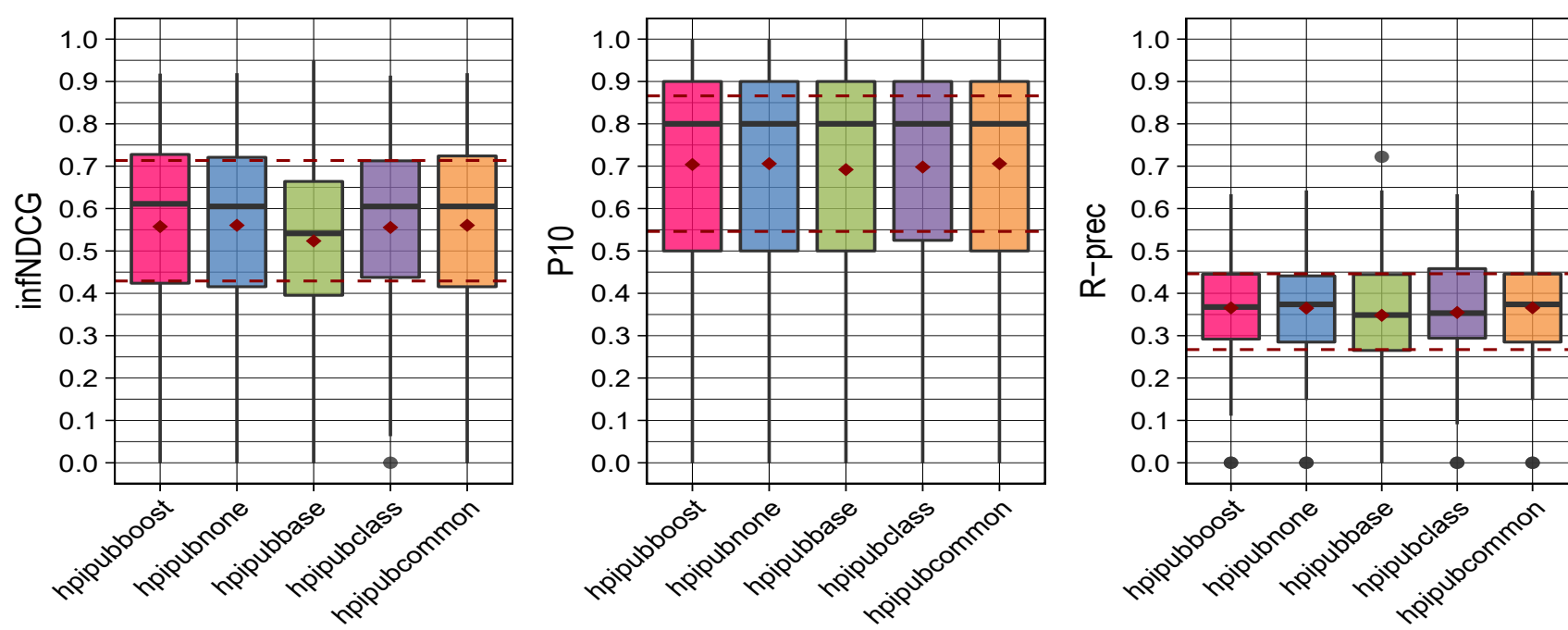


■ PM/Not PM Classifier



BIOMEDICAL ARTICLES: RUNS & RESULTS

Our best run was hpiubcommon (infNDCG of 0.5605):



Strategy	hpiubbase	hpiubnone	hpiubclass	hpiubboost	hpiubcommon
Disease Expansion + Weighting	-	Y	Y	Y	Y Except Gene Description
Keyword Boosting	Y	Y	Y	Y	Y
PM Classifier	-	-	Y	Y	-
Boost on the PM Classifier	-	-	-	Y	-
Hand-Crafted Rules	-	Y	Y	Y	Y
infNDCG	0.5235	0.5605	0.5554	0.5574	0.5605
P@10	0.6920	0.7060	0.6980	0.7040	0.7060
R-Prec	0.3481	0.3648	0.3547	0.3656	0.3658

CONCLUSIONS AND NEXT STEPS

- The difference between the worst and best run was related to gene and disease expansion,
Add support to disease hyponyms and new terminological resources
- Even though the classifier did not improve the biomedical articles results,
It could be enhanced and used to automate the keyword selection process



Michel Oleynik: michel.oleynik@stud.medunigraz.at
Erik Faessler: erik.faessler@uni-jena.de
Ariane Sasso: ariane.morassi-sasso@hpi.de

