

TUC @ CLEF IP 2012

Claims to Passage Task

What makes a baseline a baseline?

thomas.wilhelm@cs.tu-chemnitz.de

Our General Setup

Apache Lucene

Language Detection

Bag of Words

RemoveShortWordsFilter*

StopFilter

RemoveNumbersFilter*

SnowballFilter (Porter, German, French)

Our Experiments

Claims to Passages (p)

Patent and Claims to Passages (dp)

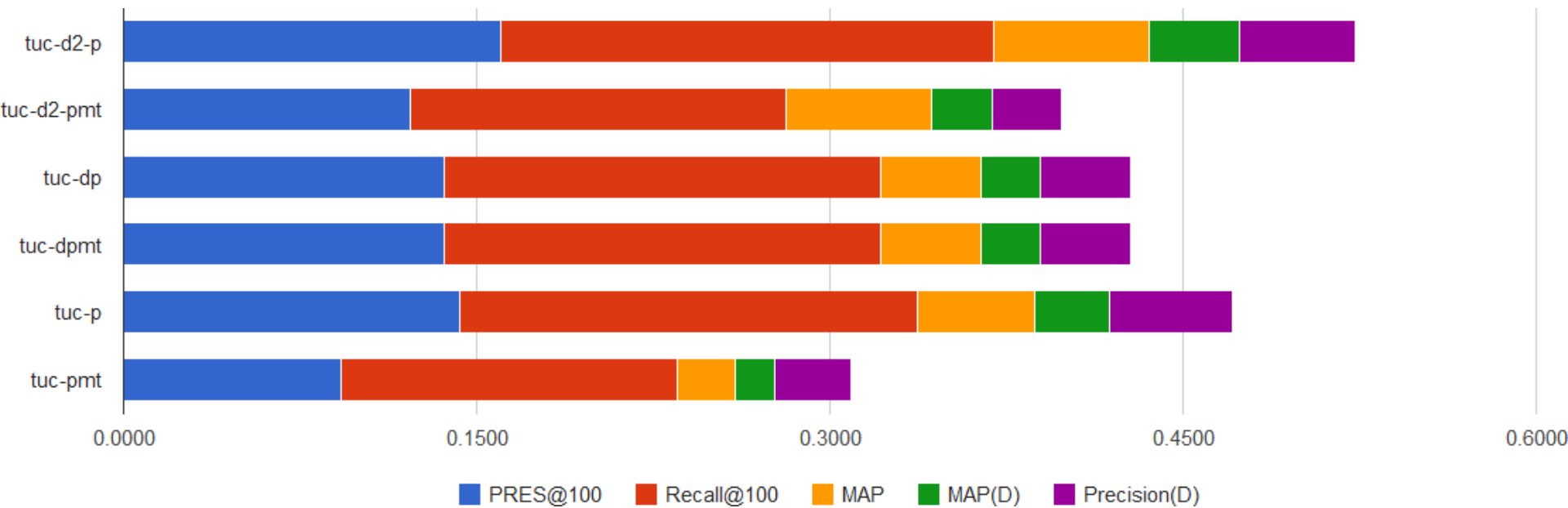
Patent to Documents, then Claims to Passages
(d2-p)

Machine Translated Claims (mt)

Claims and Documents unchanged

Our Results

EN + DE + FR



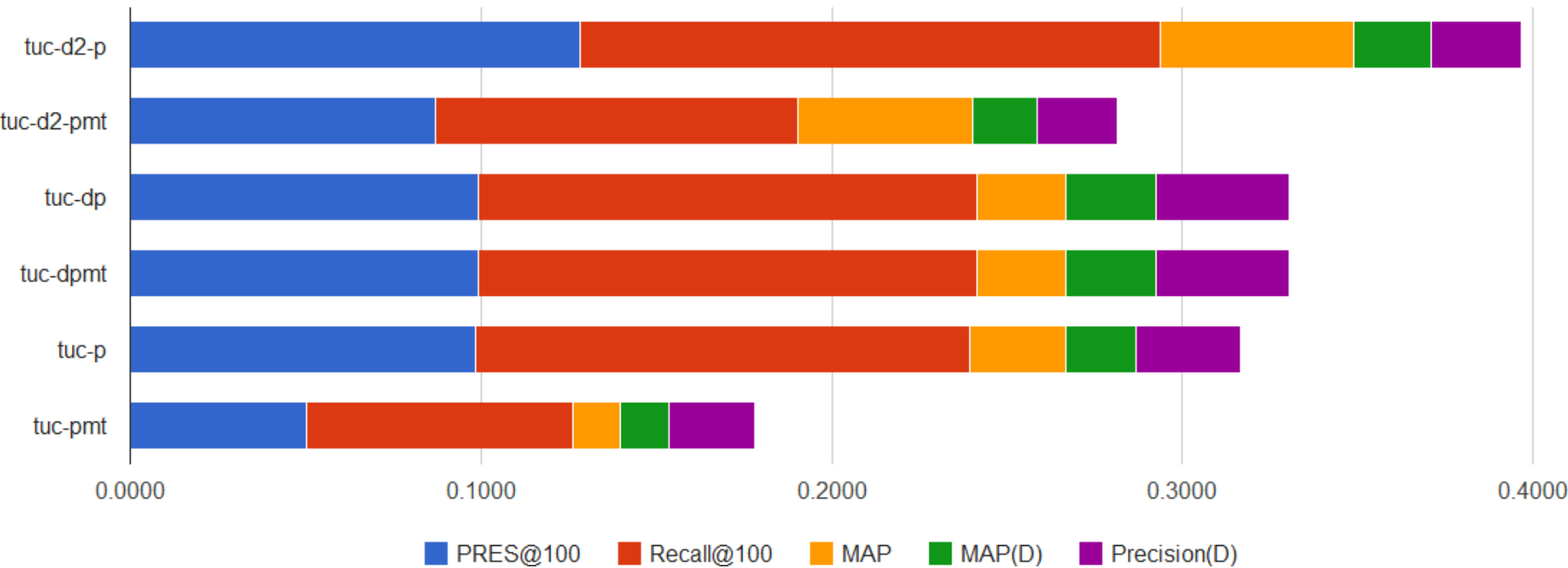
Our Results

EN



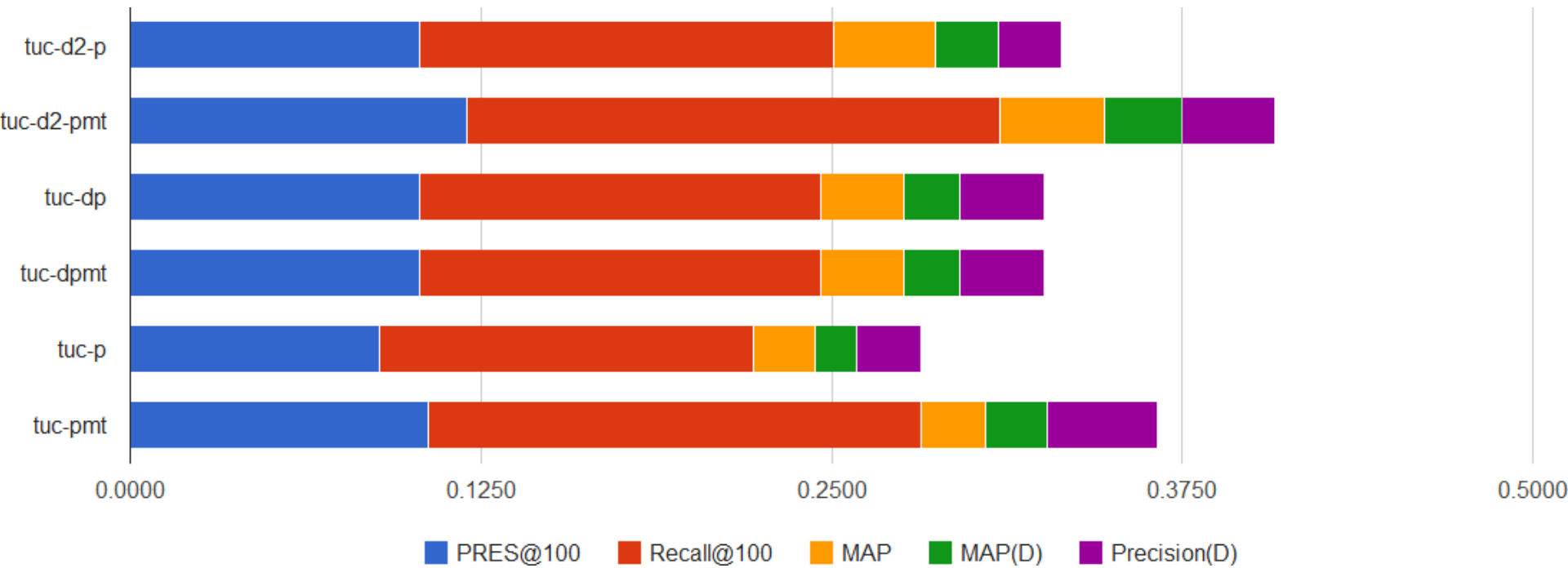
Our Results

DE



Our Results

FR



Our Results

Our baseline (out of 31 runs) is

- 9th best at PRES@100
- 7th best at MAP(D)
- 4th best at Precision(D)

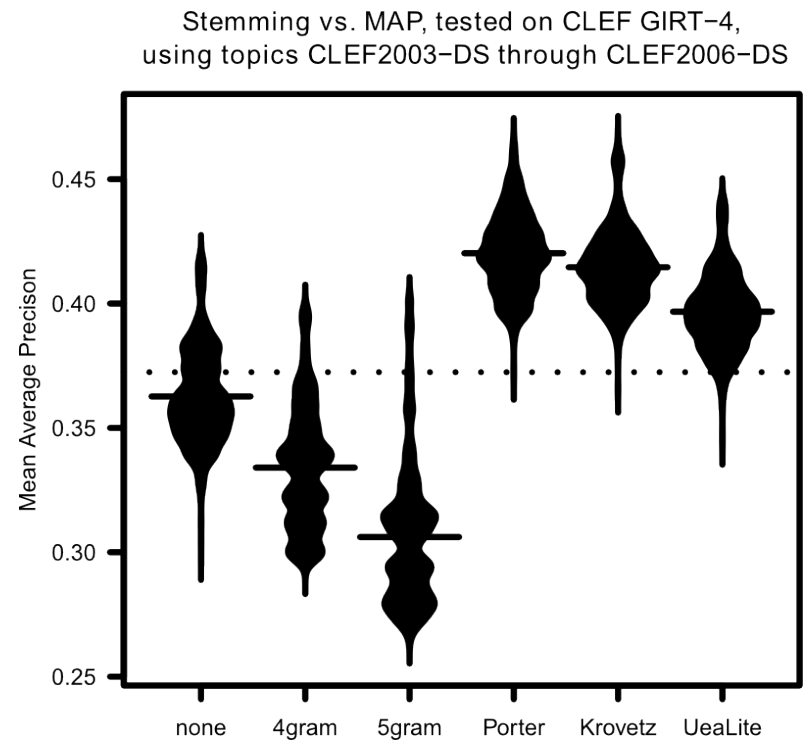
Is this still a baseline?

Lessons Learned

Translation did not work => more to investigate?

Know your test collection

Visualisation of Results is very important



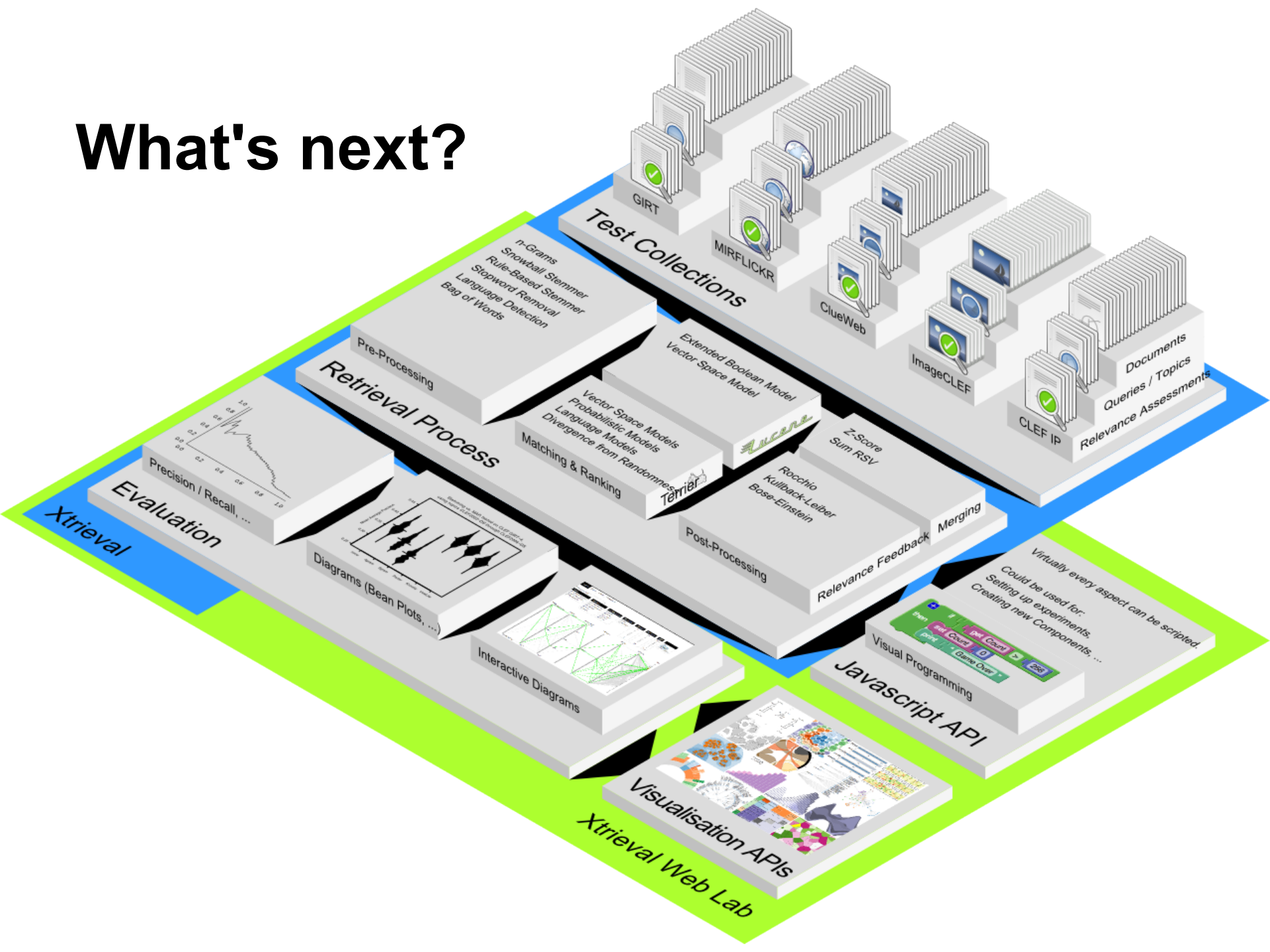
Possible Improvements

Pseudo Relevance Feedback?

Domain Specific Translations?

Terrier Ranking Models?

What's next?



TUC @ CLEF IP 2012

Claims to Passage Task

What makes a baseline a baseline?

thomas.wilhelm@cs.tu-chemnitz.de