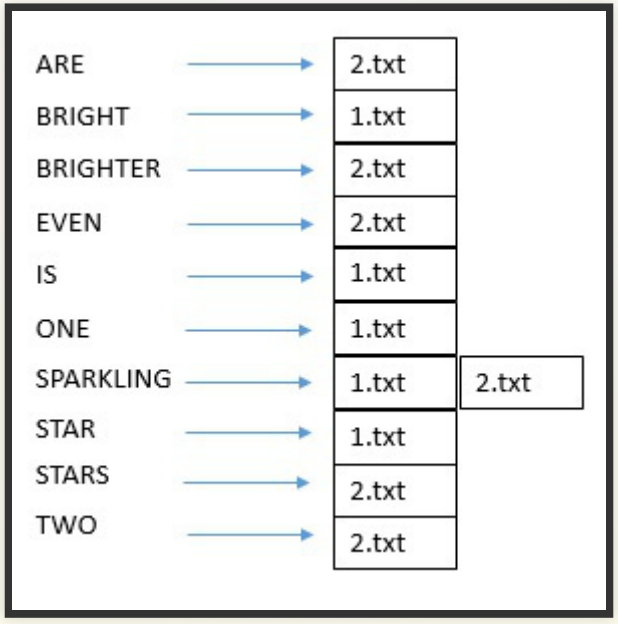
Combining spectral match with tuple-based retrieval method for more precise first-stage math-aware search

Zhong Wei

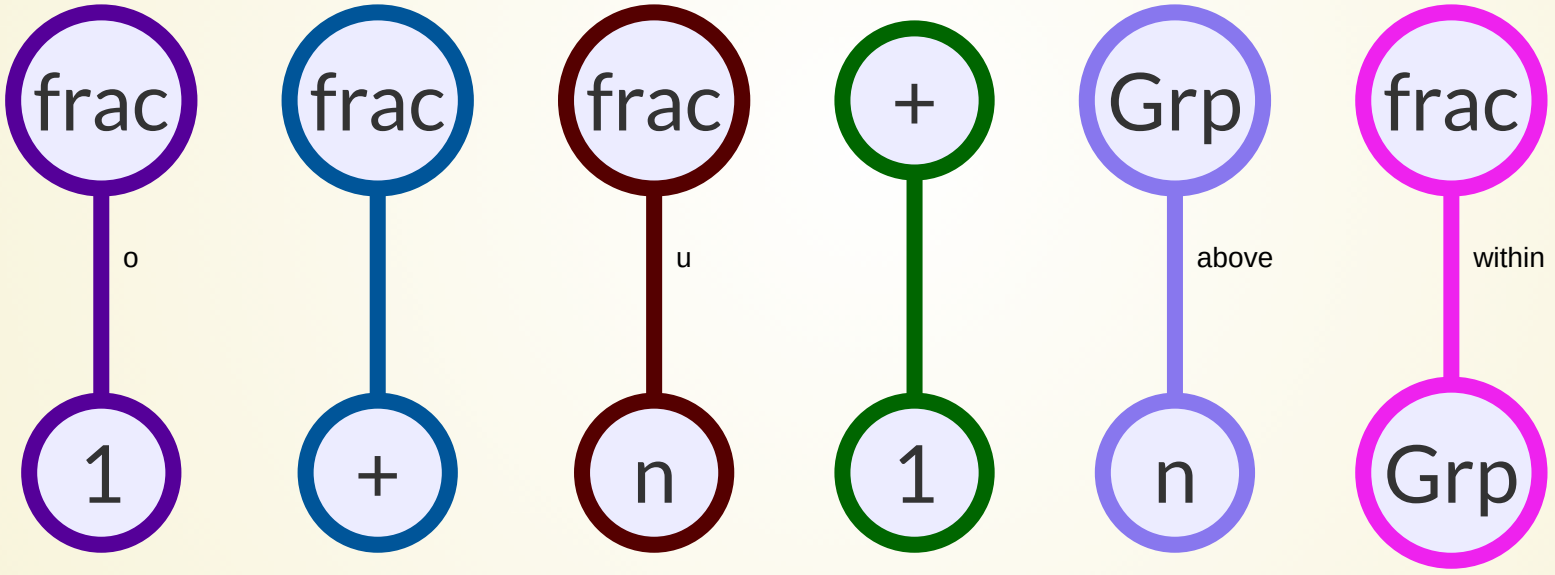
INVERTED INDEX

Term as key words to look up docID



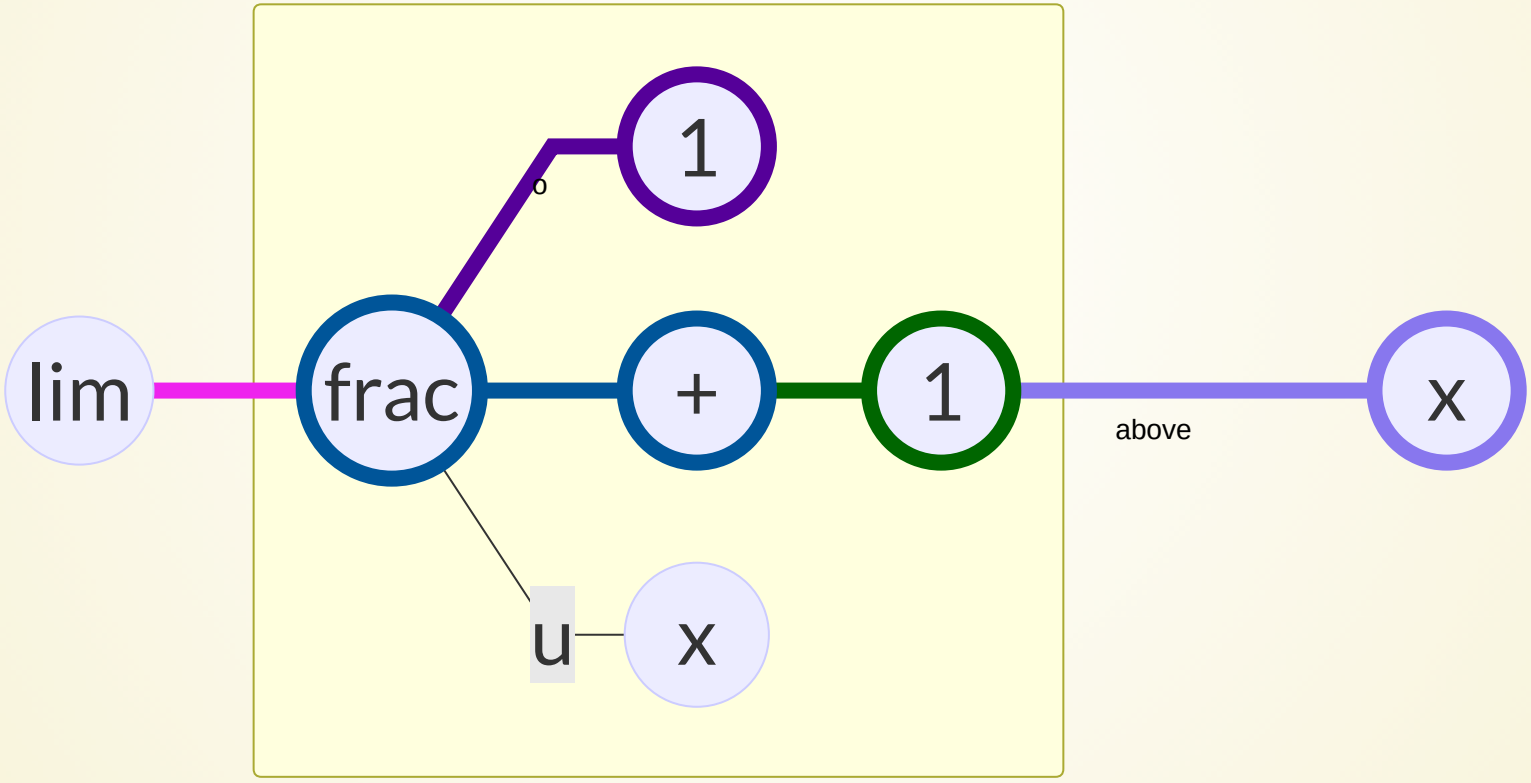
TUPLES AS INVERTED INDEX KEY WORDS

Query: $(\frac{1}{n}+1)^n$



TUPLES AS INVERTED INDEX KEY WORDS

Doc: $\lim(\frac{1}{x}+1)^x$



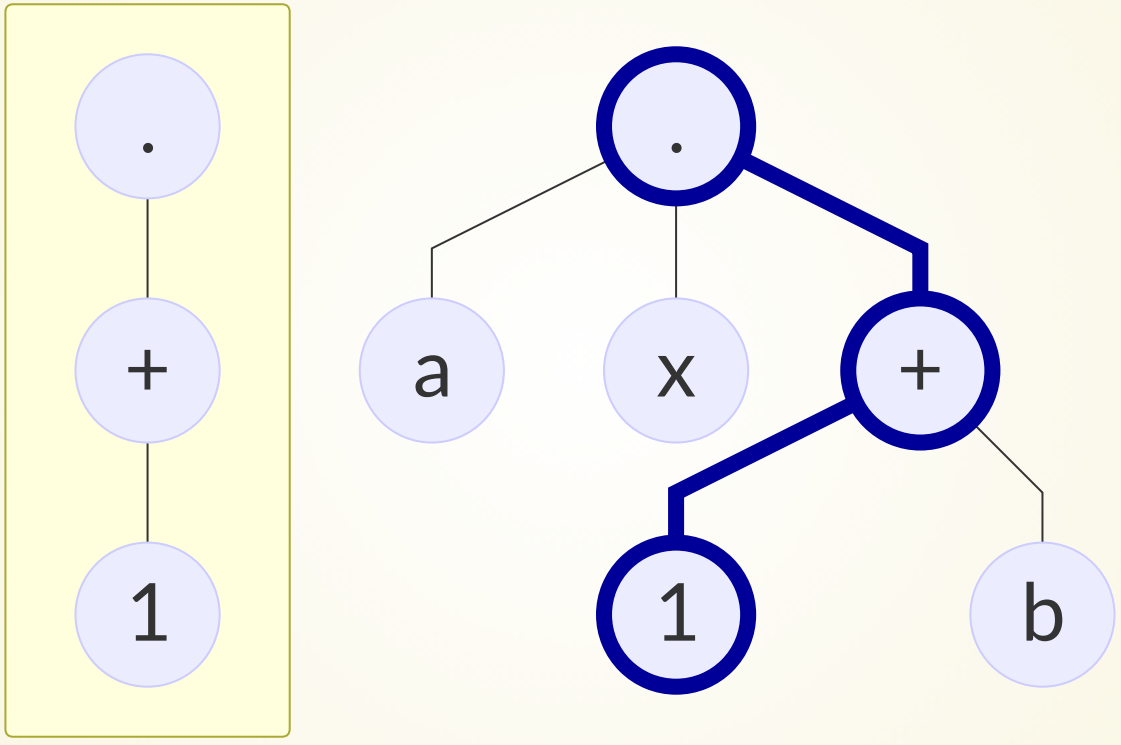
PROBLEMS

$$\frac{a}{a^2 + b} \quad \text{and} \quad \frac{a^2}{a + b}$$

are structurally different!

- High Recall, but low structural precision
- Low granularity, huge index size

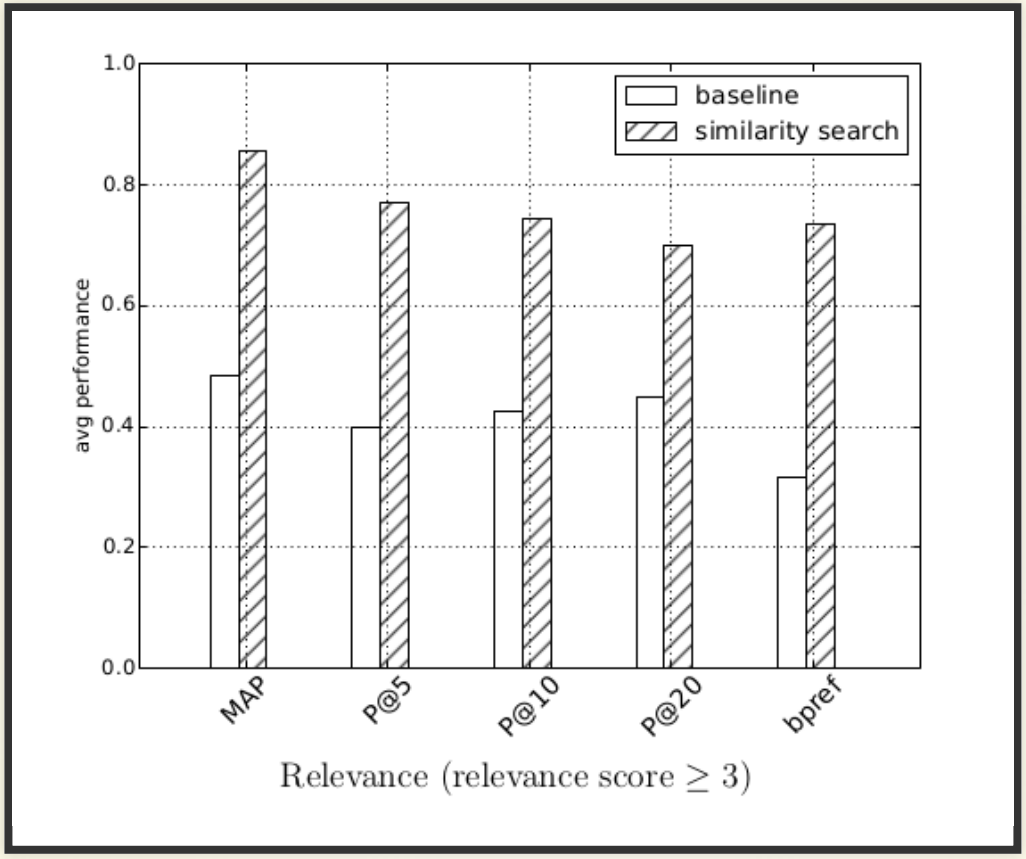
Goal: higher granularity to capture more!



METRICS OF RELEVANCE

1. 5 relevant level: Not relevant, Little relevant, Somewhat relevant, Mostly relevant, Exact
2. Precision @ top-5
3. Precision @ top-10
4. Precision @ top-20
5. $$\mathrm{MAP} = \frac{\sum_{q=1}^{Q} \mathrm{AveP(q)}}{Q}$$
6. bpref

EXPERIMENT RESULTS



Relevance compared to original system for different 5 metrics.

FUTURE WORK

1. Evaluate the impact on second-stage search.
2. Better understand tokenization effect.