Intelligent Information Systems Lab

Winter Semester 2024/2025

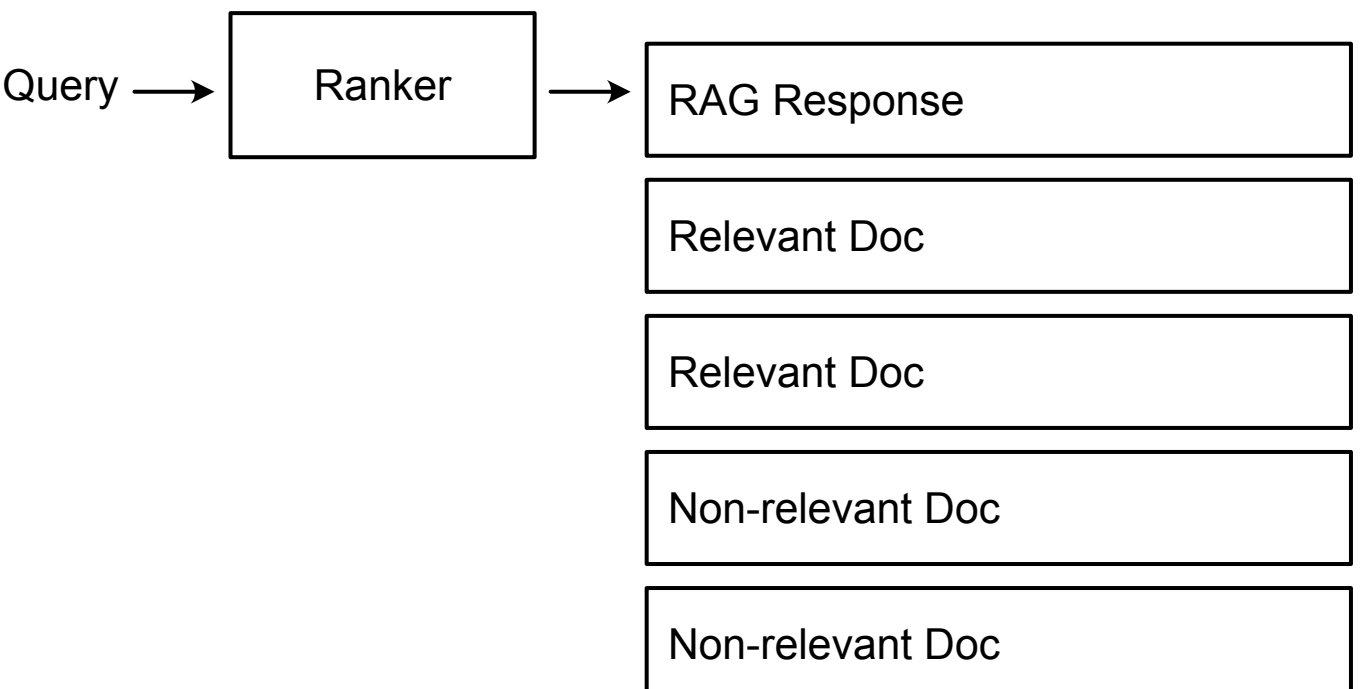
Harry Scells, Tim Hagen, and Martin Potthast

Deep Semantic Learning
University of Kassel and hessian.AI

Agenda

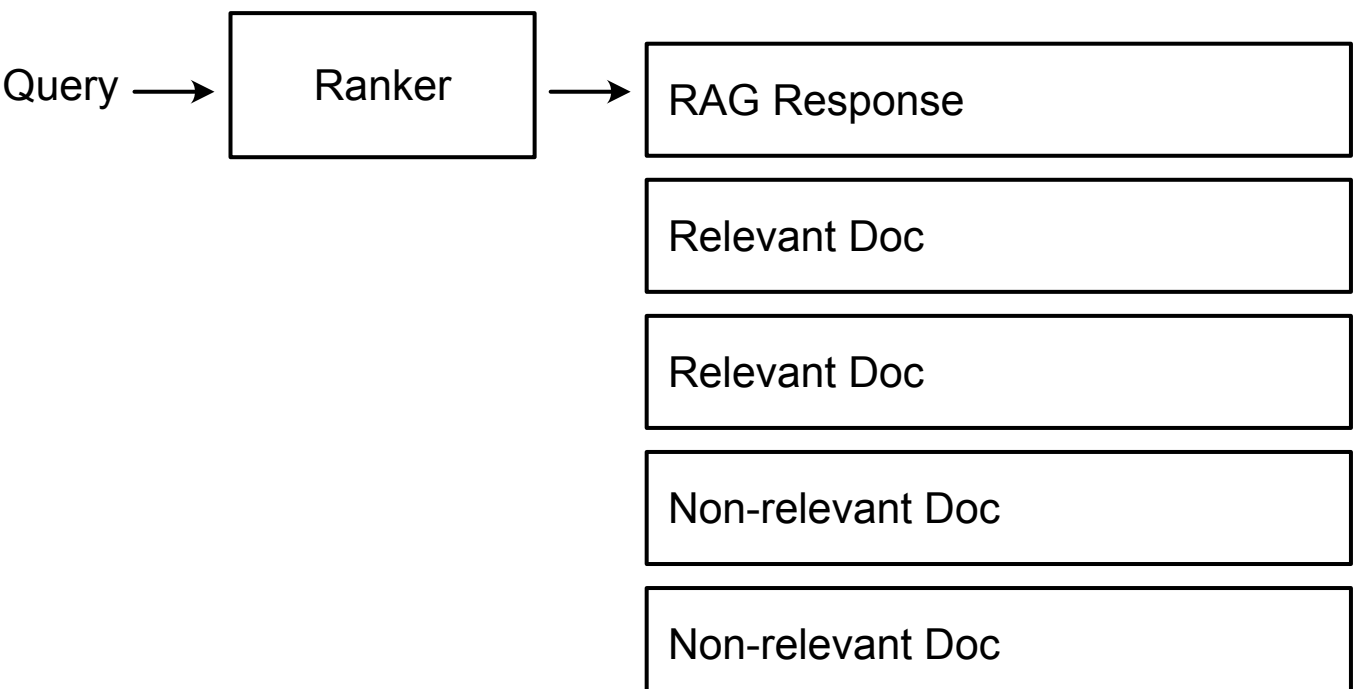
- Last week
  - Pairwise classification of RAG quality
- This week
  - Report your findings on pairwise quality classification
  - Introduction to ranking task
- Next week
  - Train and evaluate a ranking model
  - Present your findings next week, again 10-15mins

Ranking



Ranking

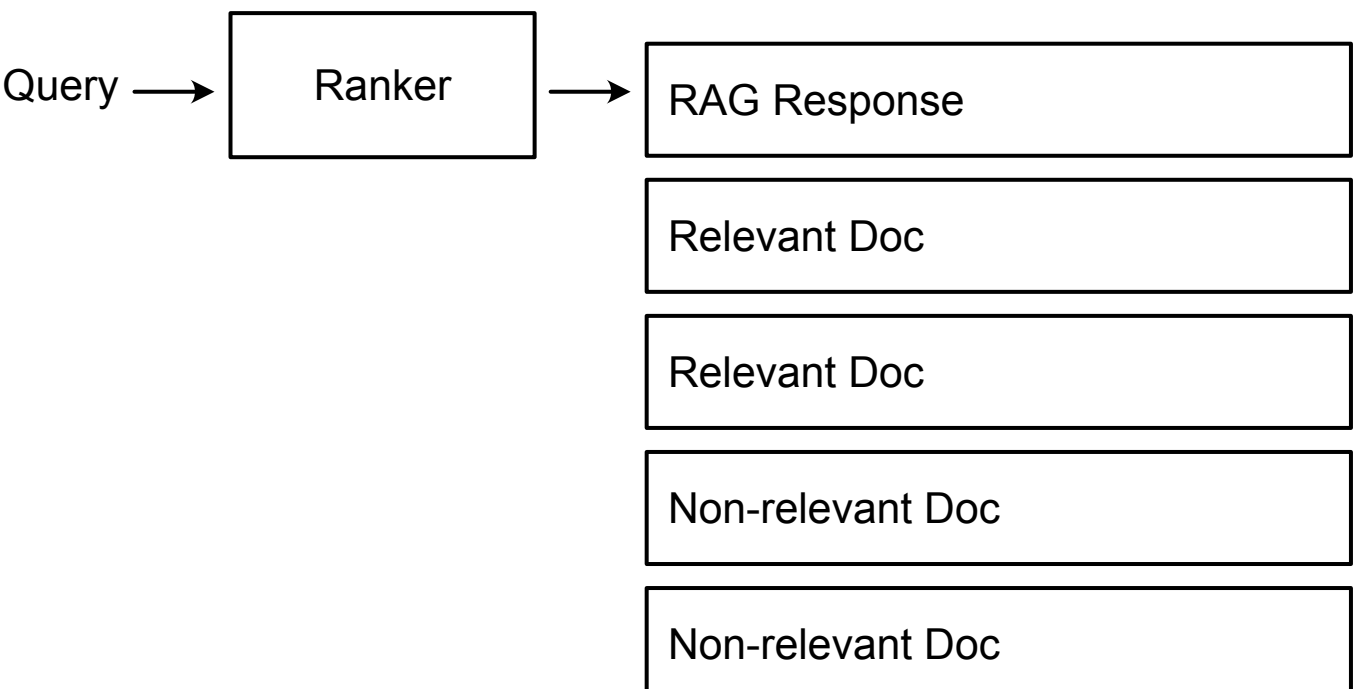
Given what you have done so far in the lab, how would you approach this task?



Ranking

Given what you have done so far in the lab, how would you approach this task?

How would you ensure that non-relevant documents would appear at the bottom?

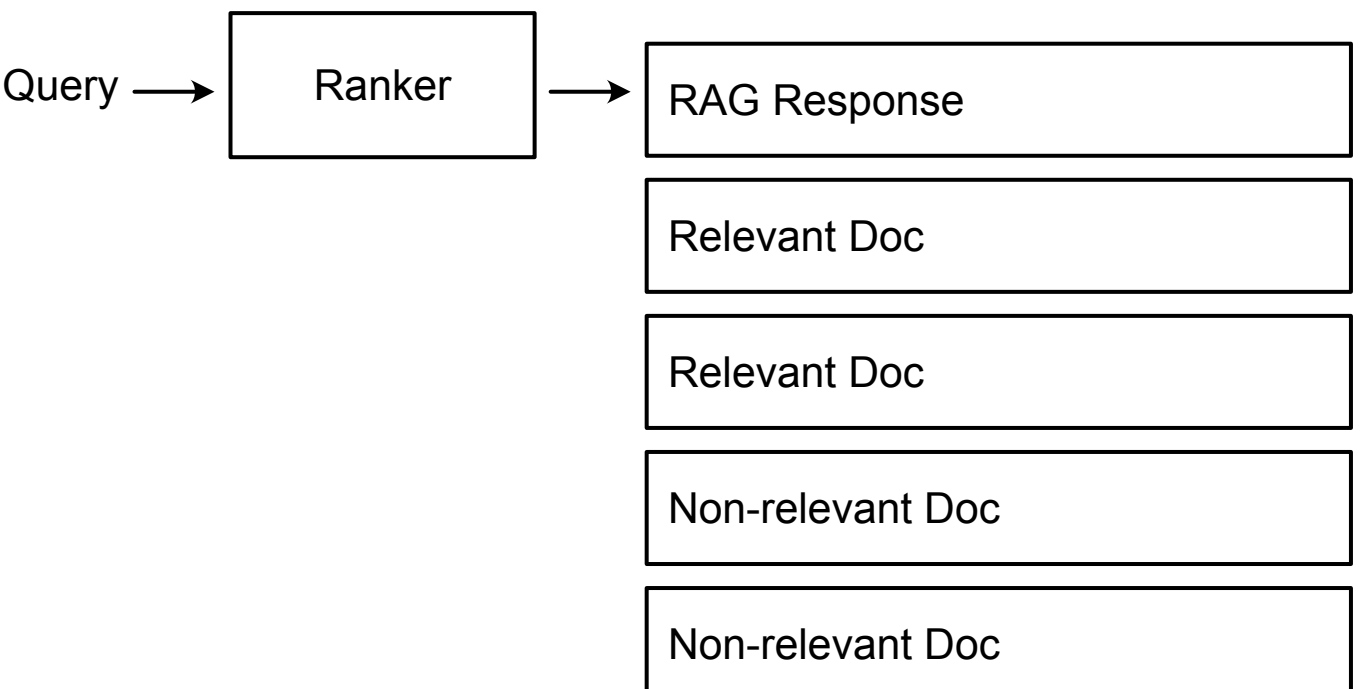


Ranking

Given what you have done so far in the lab, how would you approach this task?

How would you ensure that non-relevant documents would appear at the bottom?

How would you measure the effectiveness of this task?



Ranking Evaluation

What if we only care about the rank position of the RAG response?

Ranking Evaluation

What if we only care about the rank position of the RAG response?

➜ Reciprocal Rank

$$RR = \frac{1}{r}$$

Ranking Evaluation

What if we only care about the rank position of the RAG response?

➜ Reciprocal Rank

$$RR = \frac{1}{r}$$

What if we care about the overall ranking?

Ranking Evaluation

What if we only care about the rank position of the RAG response?

➜ Reciprocal Rank

$$RR = \frac{1}{r}$$

What if we care about the overall ranking?

➜ Discounted Cumulative Gain

$$DCG = \sum_{i}^{k} \frac{\mathsf{rel}(i)}{\log_2(i+1)}$$

Ranking Evaluation

What if we only care about the rank position of the RAG response?

➜ Reciprocal Rank

$$RR = \frac{1}{r}$$

What if we care about the overall ranking?

➜ Discounted Cumulative Gain

$$DCG = \sum_{i}^{k} \frac{\mathsf{rel}(i)}{\log_2(i+1)}$$

Any issue with this measure?

Ranking Evaluation

What if we only care about the rank position of the RAG response?

➜ Reciprocal Rank

$$RR = \frac{1}{r}$$

What if we care about the overall ranking?

➜ Discounted Cumulative Gain

$$DCG = \sum_{i}^{k} \frac{\mathsf{rel}(i)}{\log_2(i+1)}$$

Any issue with this measure?

➜ Normalise it so when averaged over multiple runs, we get a sensible value.

$$nDCG = \frac{DCG}{IDCG}$$

$IDCG$ is the 'ideal' ranking of documents, ordered from most relevant to least.