Plants

PRI - Information Processing and Retrieval (2023/2024)

T01G15
Bárbara Rodrigues - 202007163
Rubén Monteiro - 202006478
Sofia Costa - 202300565
Tiago Ribeiro - 202007589

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

INTRODUCTION

System Improvements

- **Query formulation refinement**
  - Reformulated queries
  - Improved query formulation
  - Uniformed boosts
- **Query processing**
  - Relevance feedback
  - Identified more relevant results
- **Semantic search**
  - Introduced semantic search in the system with the use of embeddings

Semantic Search

What is it?

- Ability of a system to **understand the context and intent behind a user's query**, rather than simply relying on keyword matching.
- Also known as "**dense vector search**" this method allows search systems to locate documents that share semantic relations with the query.
- Dense vectors **encapsulate semantic meaning into a confined number of dimensions**. These dimensions can be significantly fewer than the terms in a corpus, allowing a more concise representation of the document's meaning.
- A **nearest neighbor search algorithm** is used to identify the closest dense vectors for a given query.

Schema Update

add-field-type

```
{
      "name": "plantVector",
      "class": "solr.DenseVectorField",
      "vectorDimension": 384,
      "similarityFunction": "cosine",
      "knnAlgorithm": "hnsw"
    }
```

add-field

```
{
      "name": "vector",
      "type": "plantVector",
      "indexed": true,
      "stored": true
    }
```

Systems

Base System

- Without boosting

Enhanced System

- With boosting
  - Field Boosts
  - Term Boosts
  - Proximity Searches

 **Without Semantic Search**

 **With Semantic Search**

Query Formulation Refinement

Search Task #1:

Which plants can survive in cold environments?

SIMPLE

q:	(cold OR "low temperature" OR **"tolerate cold" OR "cold tolerate"**) **AND NOT** ("**damage** cold" OR "cold **damage**" OR "**not** tolerate cold" OR "**not** cold tolerate")
qf:	Introduction Description Characteristics Ecology **Cultivation**

BOOSTED

q:	(cold**^0.5** OR "low temperature"**~1^2** OR "tolerate cold"**~2^4** OR "cold tolerate"**~2^4**) AND NOT ("damage cold"**~4** OR "cold damage"**~4** OR "not tolerate cold"**~4** OR "not cold tolerate"**~4**)
qf:	Introduction Description**^2** Characteristics Ecology **Cultivation**

Query Formulation Refinement

Search Task #2:

Which plants that <u>are not trees</u> have purple flowers?

SIMPLE

q:	purple flower ~~species~~ **NOT(tree)**
qf:	Introduction Description Characteristics
q.op	AND

BOOSTED

q:	"purple flower"~5 ~~species~~ **NOT(tree)** purple^**2**
qf:	Introduction Description~~^3~~ Characteristics
q.op	AND

Query Formulation Refinement

Search Task #3:

Which plants are trees and have edible fruits <u>that are not berries</u>?

SIMPLE

q:	fruit trees edible **NOT(berry)**
qf:	Introduction **Taxonomy** Description Characteristics **Ecology Etymology Cultivation**
q.op	AND

BOOSTED

q:	fruit trees edible**^4 NOT(berry)^2**
qf:	Introduction **Taxonomy** Description Characteristics **Ecology Etymology Cultivation**
q.op	AND

OR

Query Formulation Refinement

Search Task #4:

Which plants can be used as Christmas decorations?

SIMPLE

q:	Christmas decoration winter ornamental
qf:	Name Introduction Description Etymology
q.op	OR

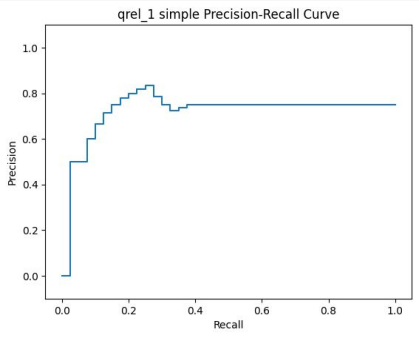
BOOSTED

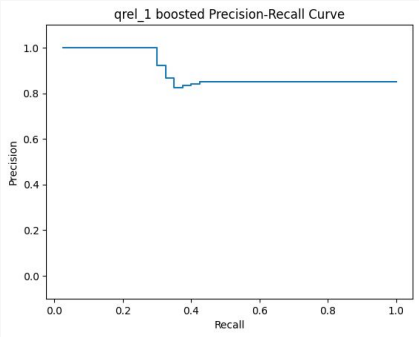
q:	Christmas^4 decoration^2 winter ornamental^2
qf:	Name Introduction Description Etymology
q.op	OR

Evaluation

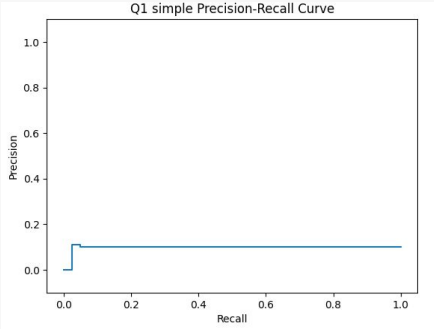
Query 1: Which plants can survive in cold environments?

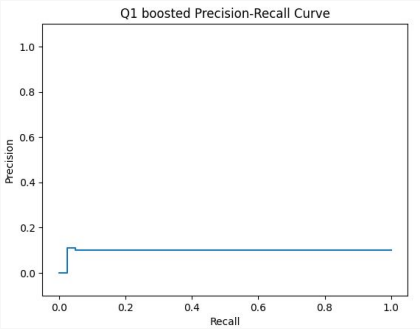
WITHOUT SEMANTIC SEARCH





WITH SEMANTIC SEARCH



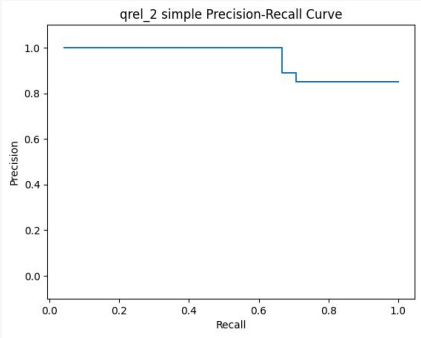


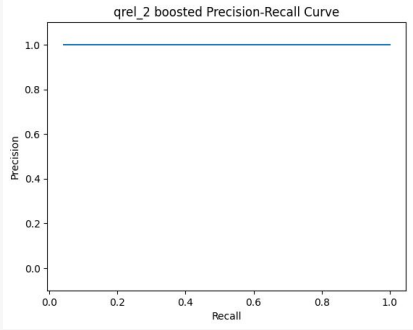
	Base System	Enhanced System	Base System with Semantic Search	Enhanced System with Semantic Search
Average precision	0.734974	0.960530	0.162500	0.266667
P@20	0.750000	0.850000	0.100000	0.100000
Individual Assessment	NRRNRRRRRRRRRNRNRNRR	RRRRRRRRRRRRNRNRNRRR	NNNNNNNRNRNNNNNNNNNN	NNRNNNNNNRNNNNNNNNNN

Evaluation

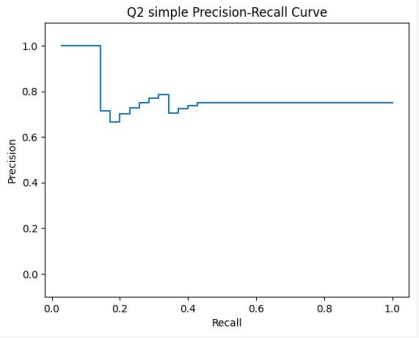
Query 2: Which plants have purple flowers?

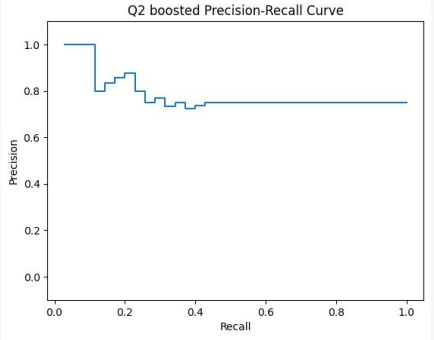
WITHOUT SEMANTIC SEARCH





WITH SEMANTIC SEARCH



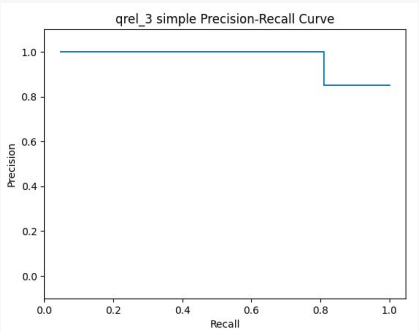


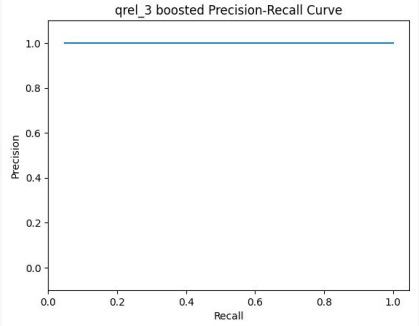
	Base System	Enhanced System	Base System with Semantic Search	Enhanced System with Semantic Search
Average precision	0.993808	1.000000	0.832752	0.855269
P@20	0.850000	1.000000	0.750000	0.750000
Individual Assessment	RRRRRRRRRRRRRRRNNRN	RRRRRRRRRRRRRRRRRRR	RRRRRNNRNRRRRRRNNRRR	RRRRNRRRRRRRRRNRRRRR

Evaluation

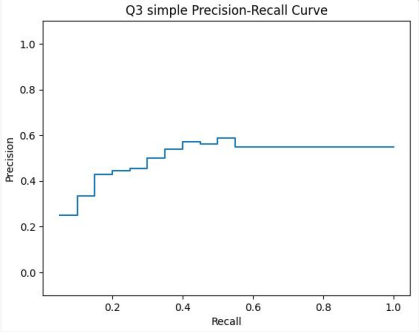
Query 3: Which plants are trees and have edible fruits that are not berries?

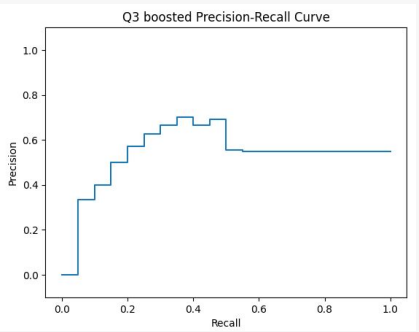
WITHOUT SEMANTIC SEARCH





WITH SEMANTIC SEARCH



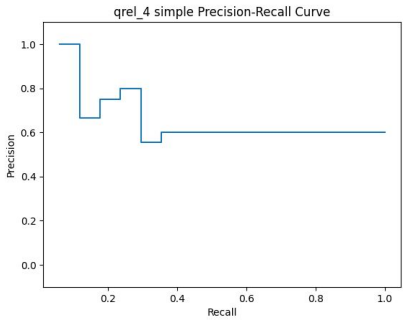


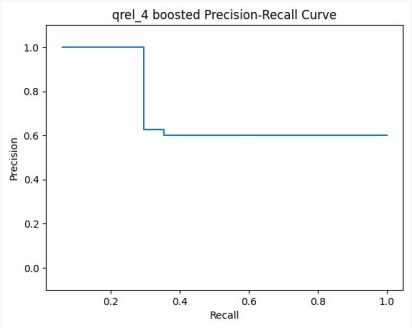
	Base System	Enhanced System	Base System with Semantic Search	Enhanced System with Semantic Search
Average precision	1.000000	1.000000	0.567073	0.615992
P@20	0.850000	1.000000	0.550000	0.550000
Individual Assessment	RRRRRRRRRRRRRRRRRNN	RRRRRRRRRRRRRRRRRRR	RNNNRNRRNRNRRRRNRRNN	NRNRNRRRRRRNRRNNNNRN

Evaluation

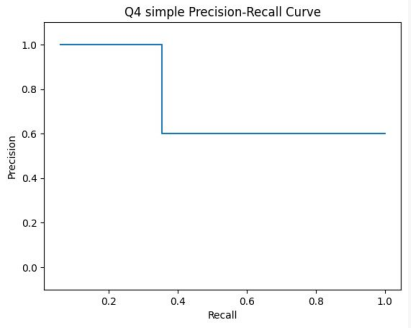
Query 4: Which plants can be used as Christmas decorations?

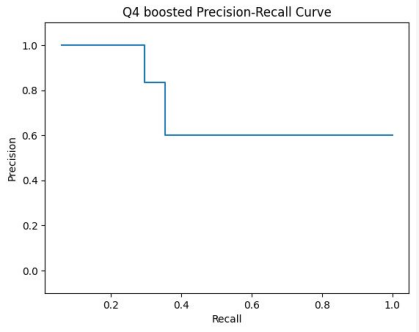
WITHOUT SEMANTIC SEARCH





WITH SEMANTIC SEARCH





	Base System	Enhanced System	Base System with Semantic Search	Enhanced System with Semantic Search
Average precision	0.775397	0.944444	1.000000	0.976190
P@10	0.500000	0.600000	0.600000	0.600000
Individual Assessment	RRNRRRNNNN	RRRRRNNNRN	RRRRRRNNNN	RRRRRRNNNN

Evaluation

Previous Base System	
Mean Average Precision	0.62

Current Base System	
Mean Average Precision	0.93

Current Base System with Semantic Search	
Mean Average Precision	0.64

Previous Enhanced System	
Mean Average Precision	0.90

Current Enhanced System	
Mean Average Precision	0.98

Current Enhanced System with Semantic Search	
Mean Average Precision	0.68

CONCLUSIONS

Thank You !

T01G15

Bárbara Rúben Sofia Tiago