

## 4.9 - Perform search reranking with semantic ranking in Azure AI Search

- [Overview](#)
  - [Introduction](#)
  - [What is semantic ranking?](#)
    - [BM25 ranking function](#)
    - [Semantic ranking](#)
      - [Semantic captions and answers](#)
    - [How semantic ranking works](#)
    - [Semantic ranking capabilities](#)
    - [Semantic ranking advantages](#)
    - [Semantic ranking limitations](#)
    - [Semantic ranking pricing](#)
  - [Set up semantic ranking](#)
    - [Configure semantic ranking](#)
  - [Exercise - Use semantic ranking on an index](#)
    - [Enable semantic ranker](#)
    - [Import a sample index](#)
    - [Configure semantic ranking](#)
  - [Knowledge Check](#)
  - [Summary](#)
- 

### Overview

Learn how to perform L2 ranking with semantic ranker in Azure AI Search.

---

### Introduction

Suppose that you work for a company that is developing an app for local tourists. You have huge amounts of information about local tourist sites, but when users search for information sometimes they don't find the most relevant articles at the top of their search results.

To improve the usability of your app, you're investigating how to return more relevant results to users by using Azure AI Search.

In this module, you'll learn how to improve search results and return more relevant results using **semantic ranking in Azure AI Search**.

After completing this module, you'll be able to:

- Describe semantic ranking
- Set up semantic ranking

- Perform semantic ranking on an index
- 

## What is semantic ranking?

Semantic ranking is a capability within Azure AI Search that aims to improve the ranking of search results. Semantic ranking improves the ranking of search results by using language understanding to more accurately match the context of the original query.

### BM25 ranking function

**Azure AI Search uses the BM25 ranking function, by default.**

The BM25 ranking function **ranks search results based on the frequency that the search term appears within a document.**

The BM25 ranking function will often achieve excellent ranking results, because a document that frequently includes a search term is often the most relevant, however, this isn't always the case.

**BM25 ranking doesn't place any relevance on the semantics of the query and ranking can sometimes be improved by adding language understanding.**

### Semantic ranking

Semantic ranking has two functions; it improves the ranking of the query results based on language understanding and it improves the response to the query by providing captions and answers in the results.

**Semantic ranking uses the BM25 ranking and calculates a new relevance score using the original BM25 ranking combined with language understanding models to extract the context and meaning of the query.**

### Semantic captions and answers

Semantic captions and answers provide additional results alongside the ranked search results that you can display to improve the understanding of the results for users.

**Semantic captions extract summary sentences from the document verbatim and highlight the most relevant text in the summary sentences.**

Semantic answers is an optional additional feature of semantic ranking that provides answers to questions. If the search query appears to be a question and the search results contains text that appears to be a relevant answer, then the semantic answer is returned.

### How semantic ranking works

Semantic ranking takes the **top 50 results from the BM25 ranking results**. The results are split into multiple fields as defined by a semantic configuration. The fields are converted into text strings and trimmed to 256 unique tokens. A token is roughly equivalent to a word in the document.

Once the strings are prepared, they are **passed to machine reading comprehension models to find the phrases and sentences that best match the query**. The results of this summarization phrase is a semantic caption and, optionally, a semantic answer.

**The semantic captions are now ranked based on the semantic relevance of the caption. The results are then returned in descending order of relevance.**

This secondary ranking uses multi-lingual, deep learning models adapted from Microsoft Bing to promote the most semantically relevant results

Note: Semantic ranker doesn't use generative AI or vectors. If you're looking for vectors and similarity search, see [Vector search in Azure AI Search](#) for details.

## Semantic ranking capabilities

See the following video for an overview of the capabilities of AI Search:

[# Azure Cognitive Search | Build semantic search into your apps | STUDIO55](#)

## Semantic ranking advantages

Semantic ranking has two key advantages over traditional search results:

- **Semantic ranking can rank results to more closely match the semantics of the original query.** This can make it more likely that the most useful documents appear at the top of the search results.
- Semantic ranking can find strings within the results to render as a caption on the search results page and to provide an answer to a question.

## Semantic ranking limitations

**Semantic ranking is applied to results returned from the BM25 ranking function.**

Although semantic ranking can re-rank the results provided by the BM25 ranking function, it will not provide any additional documents that weren't returned by the BM25 ranking function.

Semantic ranking uses the top 50 results from the BM25 ranking function. If more than 50 results are returned, only the top 50 results are considered.

For each document in the search result, the summarization model accepts up to 2,000 tokens, where a token is approximately 10 characters. Excessively long strings are trimmed to ensure the overall length meets the input requirements of the summarization step. This trimming exercise is why it's important to add fields to your semantic configuration in priority order. If you have very large documents with text-heavy fields, anything after the maximum limit is ignored.

Inputs are assembled from the "title", "keyword", and "content" fields listed in the [semantic configuration](#)

Semantic field	Token limit
"title"	128 tokens
"keywords"	128 tokens
"content"	remaining tokens

## Semantic ranking pricing

Semantic ranker is a premium feature, billed by usage. **Up to 1000 semantic ranking queries a month are available free of charge.**

For more than 1000 queries a month, you should choose standard pricing. The cost of standard pricing is based on the volume of searches, the type of searches, and the region of the search.

For more information on semantic ranking pricing, see [Azure AI Search pricing](#)

---

## Set up semantic ranking

Semantic ranking is **automatically enabled for the Azure AI Search service** at the service level and it is available for all indexes. **Semantic ranking cannot be enabled or disabled on a per-index basis.**

## Configure semantic ranking

Before you configure semantic ranking, you must have an Azure AI Search service with at least one index.

Note: For semantic ranking, the AI Search service must have a billable tier. You cannot change the pricing tier of an AI Search service. If you need another pricing tier, you will have to re-create the service.

Semantic ranking is not available in every region. Before configuring semantic ranking, check that the region of your AI Search service supports semantic ranking. To see a list of regions that support semantic ranking, see [Products available by region](#)

To choose the semantic ranking plan in the Azure portal, perform the following steps:

1. Open the Azure portal and sign in.
2. Select **All resources** and select your search service.
3. In the navigation pane, select **Settings** and select **Semantic ranker**.
4. Select the appropriate service plan. You can alter the service plan after deployment.

You can configure semantic ranking on a per-index basis. You can have multiple semantic configurations on each index.

To configure semantic ranking, follow these steps:

1. From the Azure portal home page, select **All resources** and select your search service.
2. On the navigation bar, in **Search management**, select **Indexes**.

## Search management


---

 Indexes

 Indexers

 Data sources

 Aliases

 Skillsets

 Debug sessions

3. Select your index.
  4. Select **Semantic configurations** and select **Add semantic configuration**.
  5. In **Name** type a name for your semantic configuration.
  6. In **Title field** select the field that describes the document.
  7. Under **Content fields**, in **Field name**, select a content field.
  8. Repeat the previous step for additional content fields.
  9. Under **Keyword fields**, in **Field name**, select a field with key phrases.
  10. Repeat the previous step for additional keyword fields.
  11. Select **Save**.
  12. On your index page, select **Save**.
- 

## Exercise - Use semantic ranking on an index

In this exercise, you will add semantic ranker to an index and use semantic ranker for a query.

### Enable semantic ranker

1. Open the Azure portal and sign in.
2. Select **All resources** and select your search service.
3. In the navigation pane, select **Semantic ranker**.
4. In **Availability**, in the **Free** option, select **Select Plan**.

# Semantic ranker

Semantic ranker uses deep neural networks to provide relevant results and answers based on semantics, not just lexical analysis. Additional charges may be applicable.

## Availability

<b>Free</b> 1,000 requests per month  <b>\$0.00</b> /month  <a href="#">Select Plan</a>	<b>Standard</b> First 1,000 requests per month free \$1.00 per 1,000 additional requests.  <a href="#">Select Plan</a>
--	--

## Import a sample index

1. Return to the **Overview** page of your search service.
2. Select **Import data**.



3. In **Data Source**, select **Samples**.
4. Select **hotels-sample** and select **Next: Add cognitive skills (Optional)**.
5. Select **Skip to: Customize target index**.
6. Select **Next: Create an indexer**.
7. Select **Submit**.

## Configure semantic ranking

Once you have a search index and semantic ranker enabled, you can configure semantic ranking.

You require a search client that supports preview APIs on the query request. You could use Search explorer in the Azure portal, the Postman app, the Azure SDK for .NET, or the Azure SDK for Python. In this exercise, you will use Search explorer in the Azure portal.

To configure semantic ranking, follow these steps:

1. On the navigation bar, in **Search management**, select **Indexes**.


## Search management

 Indexes

 Indexers

 Data sources

 Aliases

 Skillsets

 Debug sessions

2. Select your index.
3. Select **Semantic configurations** and select **Add semantic configuration**.
4. In **Name** type **hotels-conf**.
5. In **Title field** select **HotelName**.
6. Under **Content fields**, in **Field name**, select **Description**.
7. Repeat the previous step for the following fields:
  - **Category**
  - **Address/City**
8. Under **Keyword fields**, in **Field name**, select **Tags**.
9. Select **Save**.
10. On your index page, select **Save**.
11. Select **Search explorer**.
12. Select **View** and select **JSON view**.
13. In JSON query editor type the following text:

```
{
  "queryType": "semantic",
  "queryLanguage" : "en-us",
  "search": "all hotels near the water" ,
  "semanticConfiguration": "hotels-conf" ,
  "searchFields": "",
  "speller": "lexicon" ,
  "answers": "extractive|count-3",
  "count": true
}
```

14. Select **Search**.
15. Review the results of the query.

---

## Knowledge Check

1. How many results are returned by semantic ranking? \*

☒ Up to 50.

✓ Correct. Semantic ranking returns 50 results, or as many results as the BM25 ranking function, whichever is lower.

☐ As many results as the BM25 ranking function returns.

☐ Up to 25.

2. Which services is a prerequisite for semantic ranking? \*

☒ Azure AI Search service with a billable tier.

✓ Correct. Azure AI Search service with a billable tier is required for semantic ranking.

☐ Azure AI services with a billable tier.

☐ Azure AI Language service.

3. What are semantic captions? \*

☒ Verbatim summary sentences from the document.

✓ Semantic captions extract summary sentences from the document verbatim and highlight the most relevant text in the summary sentences.

☐ A summary of the content from the highest ranked document.

☐ A summary of the content from all documents.

---

## Summary

In this module, you learned how to:

- Describe semantic ranking
- Set up semantic ranking
- Perform semantic ranking on an index

For more information about semantic ranking, use the following links:

- **Recommended reading:** [Semantic ranking in Azure AI Search](#).
- [Enable or disable semantic ranking](#).
- [Configure semantic ranking and return captions in search results](#).