

Amazon Kendra – ML Powered Search Service

Name – Atharv Chandratre

Net ID – atharvc2

Email Address – atharvc2@illinois.edu

UIN – 657408465

Table of Contents

<i>Amazon Kendra – ML Powered Search Service.....</i>	<i>1</i>
<i>Introduction</i>	<i>1</i>
<i>What is Amazon Kendra?</i>	<i>1</i>
<i>What kinds of queries can Amazon Kendra answer?</i>	<i>1</i>
<i>What are the benefits of using Amazon Kendra?</i>	<i>2</i>
<i>How does Amazon Kendra Work?</i>	<i>2</i>
<i>Conclusion</i>	<i>3</i>
<i>References</i>	<i>4</i>

Introduction

Amazon Kendra is a machine learning-powered enterprise search engine. It enables developers to incorporate search functionality into their applications so that their customers may find information buried inside an extremely large volume of data. This data can be combined from multiple data sources, including text documents, pdfs, or data simply ingested into existing data storage solutions like AWS S3, Microsoft Sharepoint, Salesforce or Relational Databases.

What is Amazon Kendra?

With the use of complex search algorithms and natural language processing, users can search both structured and unstructured data with Amazon Kendra. Being a highly accurate and intelligent search engine, it enables fetching detailed responses for users from a large data source, which contains both structured and unstructured data. It has close integration with other AWS services like Amazon S3 and Amazon Lex. Amazon Kendra also scales extremely well to fulfill performance demands, and provides enterprise-level security while doing that. (Amazon Web Services)

What kinds of queries can Amazon Kendra answer?

Amazon Kendra can answer three major types of questions:

1. **Factoid questions**—Simple factual questions which start with the following words - who, what, when, or where. These questions have definite, fact-based answers that

can be returned in the form of a single word or phrase. The answer is retrieved from a FAQ or from your indexed documents. For example - Where is the Taj Mahal located?

2. **Descriptive questions**—Questions where the answer could be a sentence, passage, or an entire document. These questions are usually characterized by questions starting with the words – how or why. For example - How does a blockchain work?
3. **Keyword searches**—Questions where the intent and scope are not clear. For example – “Keynote Address”. As that can mean “The address of keynote” or “the keynote address given by a speaker”, Amazon Kendra has to run all the 5 stages of Natural Language Processing (lexical, syntactical, semantic, inferential and pragmatic analysis) in order to guess the user’s intent behind the query, and then try to return the most accurate response. This is handled by Machine Learning and Deep Learning models built into Amazon Kendra. (Amazon Web Services)

What are the benefits of using Amazon Kendra?

Amazon Kendra has the following benefits:

1. **Accuracy**—As Amazon Kendra uses both traditional text retrieval and machine learning, it can deliver better results than a normal search engine. Plus, it learns from the implicit feedback received from user queries, and gets better over time.
2. **Simplicity**—Amazon Kendra provides a console and API for managing your documents that you want to search. Therefore, everything is set up out of the box, and there is no need to fine tune parameters. This is especially useful when the need is to quickly implement a robust and scalable search feature onto an existing document source.
3. **Integrations**—Amazon Kendra is well integrated with the AWS ecosystem, like S3 (to ingest large volumes of data and documents), and Lex (to build chatbots from the information). It can also be integrated with 3rd party tools like Microsoft Sharepoint.
4. **User access control**— If integrated with a login system, Kendra can only show document results based on the access level of the user. For example, if used for searching within internal documents of a company, it can be configured to show different search results to an intern in the company, in comparison to the CEO of a company. (Dodel, n.d.) (Rajagopalan, n.d.)

How does Amazon Kendra Work?

The following components are present in Amazon Kendra:

- **A data source** from which Amazon Kendra takes in data. This houses all the documents and other forms of information and makes them accessible using an index. To keep your index up to date with your source repository, Amazon Kendra enables the ability to automatically synchronize a data source with an index.
- **An index** of all of the documents, that makes the data source searchable.
- **An API** for adding documents, which allows the user to directly add them to an index. Both the console and the API are available for using Amazon Kendra. Indexes can be created, modified, and removed. The API can also be used to directly make queries. (Amazon Web Services)

The user's document contents are stored in an index, which is organized internally in a way to allow the user to quickly query for information from it. Depending on the kinds of data types being used in the data source, they can be added to in a variety of ways. First of all, the user can directly connect a live data source to be indexed. This can be a document repository of some sort like an Amazon S3 bucket or a Microsoft SharePoint site. Secondly, if the user does not have a source like this, then the user can directly add documents to the index using an API. After the documents are fed into Kendra, an index can be created with one command using either the AWS CLI, AWS SDK, or the Amazon Kendra console. After indexing has completed, the user can start searching for documents. (Amazon Web Services)

Kendra can index the following kinds of documents:

- Structured text:
 - Frequently asked questions and answers (as a CSV file)
- Unstructured text:
 - HTML files
 - Microsoft PowerPoint presentations
 - Microsoft Word documents
 - Plain text documents
 - PDFs (Amazon Web Services)

It can also ingest data to index from the following data sources:

• Amazon S3 buckets • Atlassian Confluence • Custom data sources • Amazon RDS for MySQL, Amazon RDS for PostgreSQL, Amazon Aurora MySQL, Amazon Aurora PostgreSQL databases • Google Workspace Drives • Microsoft OneDrive • Salesforce • ServiceNow • Microsoft SharePoint • Amazon Kendra Web Crawler • Amazon WorkDocs • Amazon FSx • Slack • Box • Quip • Jira • GitHub • Alfresco • Zendesk • Dropbox (Amazon Web Services)

Kendra uses Natural Language Processing to understand the query, and then uses some proprietary algorithms to search the index to return the most accurate results from the document set. (Amazon Web Services)

Apart from this, Kendra also allows users to add tags to data sources, indexes and created FAQs. There are three major use cases of this. First of all, it helps identify and organize AWS resources, therefore letting you selectively get rid of data when needing to update information. Secondly, it allows you to allocate costs to each kind of tag, allowing the autoscaling to only occur until the specified limit. Finally, tagging particular indexes and data sources allows you to set access controls. For example, there might be an index which contains sensitive information. Then, you can give it a tag and set access roles to members in the organization. That way, those search results will only be returned to users who have that level of access. (Amazon Web Services)

Conclusion

Amazon Kendra is an out of the box solution to enable search capabilities to an existing data source, which uses Machine Learning to respond to queries. It is able to intelligently serve user queries by going through external data sources like S3 buckets, Microsoft Sharepoint and Relational Databases. It can be used to search through a variety of file types. Kendra can also be used to integrate with other AWS services, like using Lex to create a chat bot from

the indexed data. Under the hood, it uses three major components – data source, index and API to access both of these and perform queries.

References

- Dodel, J.-P. (n.d.). *Amazon Kendra: Transform the Way You Search and Interact with Enterprise Data Using AI* . Retrieved from YouTube:
https://www.youtube.com/watch?v=eWg9xC4vNw&ab_channel=AWSOnlineTechTalks
- Rajagopalan, K. (n.d.). *Getting Started with Amazon Kendra - Enterprise Search for Every Business - AWS Online Tech Talks* . Retrieved from YouTube:
https://www.youtube.com/watch?v=ow5DFGICHol&ab_channel=AWSOnlineTechTalks
- Amazon Web Services. (n.d.). *What is Amazon Kendra?* . Retrieved from YouTube:
https://www.youtube.com/watch?v=59kbpMnncC8&ab_channel=AmazonWebServices
- Amazon Web Services. (n.d.). *Amazon Kendra Developer Guide*. Retrieved from Amazon Web Services Documentation:
<https://docs.aws.amazon.com/kendra/latest/dg/kendra-dg.pdf>
- Amazon Web Services. (n.d.). *Index*. Retrieved from Amazon Web Services Documentation:
<https://docs.aws.amazon.com/kendra/latest/dg/hiw-index.html>
- Amazon Web Services. (n.d.). *Tags*. Retrieved from Amazon Web Services Documentation:
<https://docs.aws.amazon.com/kendra/latest/dg/tagging.html>
- Amazon Web Services. (n.d.). *Queries*. Retrieved from Amazon Web Services Documentation: <https://docs.aws.amazon.com/kendra/latest/dg/hiw-query.html>
- Amazon Web Services. (n.d.). *What is Amazon Kendra?* Retrieved from Amazon Web Services Documentation: <https://docs.aws.amazon.com/kendra/latest/dg/what-is-kendra.html>
- Amazon Web Services. (n.d.). *Data Sources*. Retrieved from Amazon Web Services Documentation: <https://docs.aws.amazon.com/kendra/latest/dg/hiw-data-source.html>
- Amazon Web Services. (n.d.). *Types of Documents*. Retrieved from Amazon Web Services Documentation: <https://docs.aws.amazon.com/kendra/latest/dg/index-document-types.html>
- Amazon Web Services. (n.d.). *Documents*. Retrieved from Amazon Web Services Documentation: <https://docs.aws.amazon.com/kendra/latest/dg/hiw-documents.html>