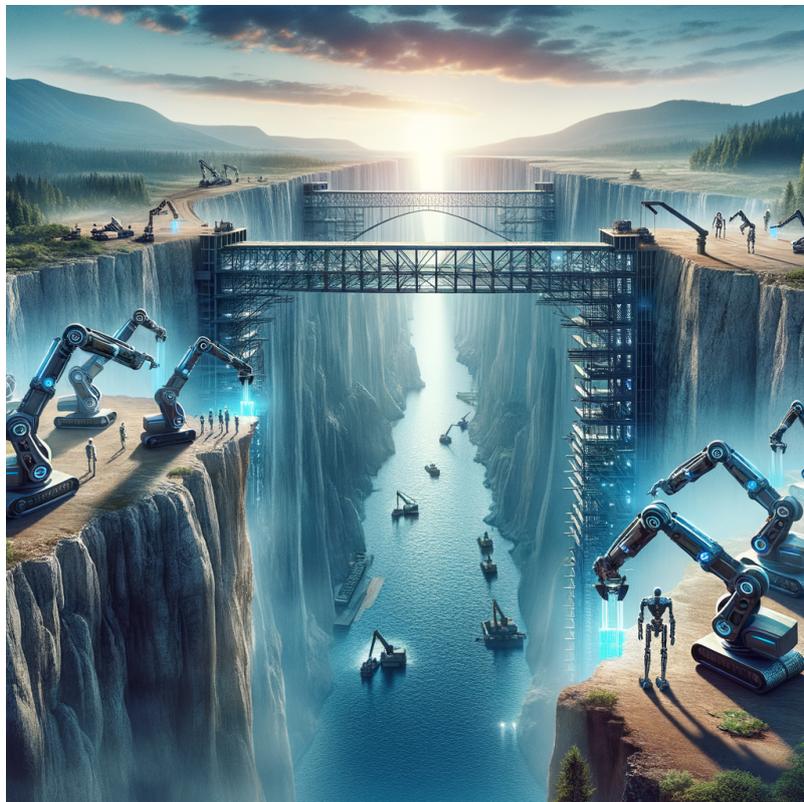


Bridging the Gap

Comparing Text Retrieval Systems for B2B Trading

Computer Science Bachelor Thesis

Windisch, March 2024



Students George Rowlands
 Raphael Lüthy

Client Nitish Patkar

Supervisors Norbert Seyff
 Nitish Patkar

Project Number 23HS_IIT36

Fachhochschule Nordwestschweiz, Hochschule für Technik

Abstract

Through our analysis and observations, it is evident that B2B trading platforms have not kept up with recent trends in searching. They require users to use precise keywords in their searches to yield satisfactory results rather than allowing them to use natural language, thus hindering the identification of potential business partners. Furthermore, these platforms offer minimal consulting or support for finalizing trading transactions once a suitable partner is identified; opening up a potential gap in the market.

We aim to bridge this gap by laying the foundations for a platform that facilitates international business collaboration. Focusing on implementing and comparing different approaches to searching with three information retrieval systems: PostgreSQL Full Text Search, semantic search and Elasticsearch. These three information retrieval systems leverage a unique dataset synthetically generated by GPT-4 to facilitate the search of companies, the products they have to offer, and the needs they want to have fulfilled. Using a test dataset of queries and expected results, we meticulously evaluate each system, analyzing factors such as latency, search accuracy and a unique approach that assesses the overall relevance of results using Large Language Models (LLMs) as judges. In our work, we find that whilst semantic search has great potential with its capabilities of understanding a query's contextual meaning, without enhancing it with additional techniques, it can not outperform the current industry standard search solution, Elasticsearch.

Keywords:

B2B Trading, Information Retrieval, Large Language Models, Text Embeddings, Microservice Platform

Acknowledgements

We would like to express our sincere gratitude to Prof. Dr. Norbert Seyff and Dr. Nitish Patkar for their invaluable professional support throughout our entire thesis. Their guidance and expertise were instrumental in shaping this research and ensuring its quality.

We are also grateful to Dr. Fernando Benites and Claudio Paonessa. They not only validated our approaches but also assisted us with their expertise on natural language processing subjects, which significantly improved the strength and direction of our research.

We extend our thanks to Jacqueline Rowlands for her meticulous proofreading, which helped us ensure the clarity and accuracy of our work.

Finally, we would like to express our deepest appreciation to all the participants who took part in our survey. Their valuable feedback provided essential data and helped us evaluate the effectiveness of our search approaches.

Contents

1	Introduction	1
1.1	The Gap in the Market	1
1.2	Our Bridge	2
2	Principles of Information Retrieval	4
3	State of the Art	6
3.1	B2B Trading Platforms	6
3.2	Text Embeddings	9
3.3	Information Retrieval	11
4	Our Approaches to Searching	13
4.1	PostgreSQL Full Text Search	13
4.2	Semantic Search with pgvector	14
4.3	Elasticsearch	14
5	Our Dataset	15
5.1	Existing Datasets	15
5.2	Scraping Data from other Platforms	15
5.3	Creating our own Dataset	16
5.3.1	Prompting Chain	18
5.3.2	Data Analysis	19
5.3.3	Data Enhancements	19
6	Conceptual Solution	20
6.1	Data Models	20
6.2	Frameworks and Programming Languages	21
6.3	Monolith vs Microservice Architecture	23
6.4	User Interface Design	26
7	Solution	28
7.1	Frontend	28
7.1.1	Login Page	28
7.1.2	Search Page	29
7.1.3	Multi-Language Support	30

7.2	Backend	31
7.3	Implementation of our Search Approaches	32
7.3.1	Changing between Systems	32
7.3.2	PostgreSQL Full Text Search	34
7.3.3	Semantic Search	35
7.3.4	Elasticsearch	40
7.4	Code Testing	42
7.4.1	Linting	42
7.4.2	Frontend	42
7.4.3	Microservices	43
8	Evaluation	45
8.1	Latency Evaluation	45
8.1.1	PostgreSQL FTS	47
8.1.2	Semantic Search with HNSW Index	48
8.1.3	Semantic Search without HNSW Index	49
8.1.4	Elasticsearch	50
8.1.5	Findings	51
8.2	Human Evaluation	52
8.2.1	Participants	54
8.2.2	Comparison of Search Results	56
8.2.3	Comparison with other Search Tools	57
8.2.4	Findings	58
8.3	Automated Evaluation	59
8.3.1	Test Dataset	59
8.3.2	Metrics	59
8.4	Human Queries vs LLM Queries	62
8.4.1	LLMs as Judges	63
8.4.2	Findings	65
8.5	Impact of Data Quality	66
8.5.1	Findings	66

9 Conclusion and Future Work	67
9.1 Conclusion	67
9.2 Learnings	68
9.3 Future Work	69
9.3.1 Matchmaking	69
9.3.2 Reranker	69
9.3.3 Query Transformations	70
9.3.4 Hybrid Search	70
9.3.5 Usability & Trustworthiness	70
9.3.6 Production	71
Declaration of Honesty	72
References	73
List of Figures	78
List of Tables	80
List of Listings	81
A Prompts Appendix	82
B Onboarding Wireframes Appendix	84
C Survey Appendix	87

Glossary

- API** An Application Programming Interface (API) is a defined communication protocol that allows other software applications to interact with the API system. This term is most often used to refer to the backends of web applications.
- CRUD** CRUD stands for the Create, Read, Update and Delete operations. It is a common acronym used when discussing operations on persistent storage.
- CSS** Cascading Style Sheets (CSS) is a programming language for applying styles to web pages. It controls how web elements are displayed, including font, color, spacing, layout and animations.
- Docker** Docker is a tool for developers to package applications with all their dependencies into a standardized unit called a “container”. These containers can then be easily deployed and run on any system with Docker installed, ensuring consistent performance and behaviour regardless of the environment.
- DoS** Denial of Service attacks (DoS) are a method of attacking systems by creating a large artificial amount of requests. The goal of these attacks is to either slow down the system or to take it offline completely.
- gRPC** gRPC is a modern open-source high-performance Remote Procedure Call (RPC) framework that can run in any environment. It can efficiently connect services in and across data centers with pluggable support for load balancing, tracing, health checking and authentication. It is also applicable in the last mile of distributed computing to connect devices, mobile applications, and browsers to backend services.
- IR** Information Retrieval (IR) refers to the process of searching for specific information from a large collection of data, often referred to as “documents”. This process forms the basis of most search engine operations.
- IRS** Information Retrieval Systems (IRS) are systems that solve the task of searching for documents based on a search query using IR.
- JSON** JavaScript Object Notation (JSON) is a lightweight and human-readable format for storing data. It consists of key-value pairs and has the ability to display basic data structures like arrays and objects.
- lexeme** A lexeme is a term that has been “normalized” so that different forms of the same word can be treated equally. This normalization typically includes transforming upper-case letters to lower-case and often involves removing suffixes, such as “s” or “es”.
- LLM** Large Language Models (LLMs) are neural networks that have been trained on large corpuses of text data. This allows them to understand natural language and generate human-like text in response to a wide range of prompts.
- NLP** Natural Language Processing (NLP) is a subfield of computer science and focuses on the interaction between human language and computers.
- OAuth** OAuth is a system that allows users to grant access to their accounts on one site (like Facebook or Google) to another application, without sharing their passwords. This keeps user data secure and simplifies login processes.
- protobuf** Protocol Buffers (Protobuf) is a method of serializing structured data. It is used to develop programs that communicate with each other over a wire or for storing data. In this project, we use it together with gRPC.
- REST** Representational State Transfer (REST), is an architectural style for designing APIs. This standardized approach makes them easy to develop, understand, and use across different applications and platforms.
- stub** A stub is a generated adapter for gRPC, which is generated from the protobuf file to be used in applications.

Swagger Swagger is a suite of tools built around the OpenAPI Specification, used for designing, documenting, and interacting with APIs.

SwitchEngines SwitchEngines is a service offered by SWITCH, a Swiss research and education network, that provides on-demand virtual machines with computing and storage power for research and education purposes.

tsquery The `tsquery` type represents a text search query in the form of a sorted list of distinct lexemes.

tsvector The `tsvector` type represents a document in a form-optimized for text search. It is a sorted list of distinct lexemes.

UI User interface (UI) refers to the buttons, menus, and other elements on a computer or device with which users interact.

Vector Vectors are used in many fields, including computer science and mathematics. Depending on their use case, they can represent a list of numbers or a point in a multidimensional space.

1 Introduction

In this thesis, we describe the development of a platform that has the potential to improve business-to-business (B2B) trading. This platform not only aims to connect businesses but also uses cutting-edge search capabilities to unlock hidden trading opportunities. Specifically, we focus on developing and comparing three different approaches to searching, as this is the key to exploring a trading platform and identifying trade partners.

1.1 The Gap in the Market

Two elements need to be considered: The market for B2B trade itself and the gap within that market.

In recent years, India has emerged as a market with substantial growth prospects. Notably, it has witnessed the first Indian IT company to achieve a market capitalization exceeding \$100 billion with Infosys [1]. India has the largest population globally, offering a vast array of skills and a large pool of potential buyers and sellers [2]. Furthermore, the Indian government actively supports the growth of B2B trading through initiatives such as the Open Network for Digital Commerce (ONDC) ¹ and the Open Credit Enablement Network (OCEN).² These initiatives aim to streamline B2B transactions and foster a secure environment for businesses to develop.

Switzerland's strengths lie in its political and economic stability. Even after the collapse and takeover of its second-biggest bank, Credit Suisse, by UBS, the economy remains stable [3]. Switzerland boasts a long-standing reputation for quality and reliability. This is further solidified by its proven track record in innovation [4], evident in its global leadership across various sectors like pharmaceuticals, medical technology, finance, and engineering [5]. A vibrant startup ecosystem, fostered by theory-oriented institutions like the Eidgenössische Technische Hochschule Zürich (ETH) and practice-oriented institutions like the Fachhochschule Nordwestschweiz (FHNW), cultivates a diverse but balanced and skilled workforce that, coupled with a strong emphasis on research and development, ensures Switzerland's continued position at the forefront of innovation.

These two countries strongly complement each other, offering exciting possibilities for international B2B trading, particularly in the IT, engineering, and manufacturing sectors. However, the potential is not limited to this specific pairing of countries. Similar opportunities exist for other countries with significant growth prospects.

In March 2024, Switzerland and the other members of the European Free Trade Association (EFTA) agreed to a free trade deal with India after 16 years of negotiations [6]. This agreement is estimated to be worth 88.5 billion CHF and creates a golden opportunity for our trading platform to be the first to bridge the market gap between these two powerhouses.

Over the last few years, the way we search for information on applications like Google has undergone significant changes. Initially, keywords were matched between the search and result texts. However, this evolved to include more advanced techniques such as sub-string and token-based searching, graph-based and machine-learning-based solutions, and most recently, vector-based semantic search. These more recent developments can be accredited to the popularity and advancements of artificial intelligence, particularly in the field of natural language processing [7]. Especially with the release of the transformer architecture [8], there has been a significant increase in research and open-sourcing of artificial intelligence models. Semantic search has profited the most from these advancements as it is based on text embeddings [9]. Research into popular large language models (LLMs) [10] and the foundational transformer architecture progresses in tandem with advancements in embedding models. This is because embedding models act as the initial processing step for input data within the transformer architecture.

¹<https://ondc.org/>

²<https://ocen.dev/>

After analyzing other platforms for B2B trading, we found that they have not kept up with the recent trends in searching. Most platforms do not offer accurate search results for natural language queries and only offer exploration at the company level, not the product level. This means that users can only search for companies; they can not search for products directly. This results in users having to manually go a step further after identifying a promising company to explore their products and does not allow users to easily compare the different companies and products. Most platforms also do not have any automatic matchmaking capabilities, such as recommendations for companies to find suitable business partners based on their search history. Furthermore, within the platforms, there is no possibility of directly hiring a consultant to finalize a trading agreement when the companies have found a suitable product to buy or sell. Instead, they leave the daunting challenge of navigating the complicated regulations and laws of the relevant countries to the companies.

Hence, Richard Elsasser, a business student at FHNW, wrote a management summary as part of his bachelor thesis [11] in the spring semester of 2023 under the supervision of Professor Pieter Perrett to analyse the necessity for a platform that facilitates international business collaboration. The thesis aimed to discuss the potential challenges associated with B2B trading and the opportunities for such a platform. The management summary concluded that there is a potential market for such a platform. We have therefore chosen to explore this conclusion in our work here and look into the development of such a platform.

1.2 Our Bridge

We aim to lay the foundations for a platform for international B2B trading to bridge the gap in the market mentioned above. Understanding that efficient and insightful exploration is the lifeblood of a successful B2B trading platform, we embark on a comprehensive evaluation to compare and contrast the technical and perceived differences between search results provided by three distinct search solutions with the aim to find the best fit for our B2B trading platform. The three search solutions we evaluate and compare are as follows:

- Our baseline is PostgreSQL’s built-in Full Text Search.³
- Semantic search using the highly efficient and accurate embedding model, `all-MiniLM-L6-v2`⁴ and the `pgvector` extension⁵ for PostgreSQL.
- The current industry standard, Elasticsearch.⁶

In addition to these three search solutions, we build a user interface which is independent of the underlying search solution and allows users to have a first impression of the new platform. This user interface encompasses basic user management, offers multi-language support and showcases our search capabilities for potential future users or investors.

We form the basis of our research work around the following two key research questions:

Question 1: How does our platform compare to similar platforms?

In order to answer this research question, the following three sub-questions must be answered:

- **RQ1.1** What features can aid a user during exploration?
- **RQ1.2** How do B2B trading platforms’ search features compare?
- **RQ1.3** How do users perceive the search results of our platform compared to those of similar platforms?

³<https://www.postgresql.org/docs/current/textsearch.html>

⁴<https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>

⁵<https://github.com/pgvector/pgvector>

⁶<https://www.elastic.co/enterprise-search>

Question 2: How do different Information Retrieval Systems (IRS) for exploring companies and their products compare?

To answer this question, we define and answer the following three sub-questions:

- **RQ2.1** Which of our implemented IRS provides the most relevant results when exploring companies and their products?
- **RQ2.2** How do the response times of the various IRS compare?
- **RQ2.3** How do the different IRS behave when subjected to diverse qualities of data?

In summary, our contributions are:

- Comparisons of different information retrieval systems for the business case of exploring companies and products.
- The initial groundwork for the platform, including the architectural design, user authentication and authorization, a multi-language user interface for exploration and a setup for end-to-end microservices testing.
- A new synthetic dataset containing companies, products and needs.
- The dataset also includes a test dataset for evaluation containing pairs of queries and expected results. The queries are a combination of synthetically generated and manually written queries from a given product or company description. Combining manually created and synthetic queries leverages the strengths of each approach. Synthetic queries offer scalability and efficiency, while manually created queries capture genuine user behaviour and search patterns.
- A novel way to evaluate the relevance of a search result for a given query using large language models as judges. These models are particularly well-suited for this task due to their deep understanding of language and being highly efficient and cost-effective compared to human evaluation.

Our work is structured as follows: We first provide context on information retrieval (section 2) and the three approaches we will explore to provide the platform with relevant search results (section 4). Then, we review the current state of B2B trading platforms and state-of-the-art information retrieval systems (section 3). Following this, the focus will be on our own work, detailing the data used in the platform and how we generated it (section 5). We discuss and explain our proposed solution, including the platform's architecture and why we chose specific frameworks (section 6). The heart of our work then lies in implementing and evaluating our search solutions (section 7 and 8). Here, we meticulously assess various aspects of each information retrieval system, including latency, accuracy and overall relevance of results. Finally, we reflect on the work we carried out and present possible future work (section 9).

2 Principles of Information Retrieval

The aim of this section is to discuss general concepts and techniques of information retrieval and how they relate to our work.

Information retrieval systems are structured frameworks for locating and accessing relevant information within a collection of data. They employ various techniques to match user queries with document representations based on content and relevance.

Some examples of information retrieval are:

- Searching for information in a document *e.g.* searching for a word in a dictionary or book.
- Searching for documents themselves *e.g.* searching for a book in a library or a bookshop.
- Searching for the metadata that describes data *e.g.* searching for a book in a library by its author.

In information retrieval, search targets are abstracted to documents. Therefore, the task of retrieving documents is commonly referred to as **document retrieval**. If the retrieval process only uses the textual content within these documents, the task can be specified as **text retrieval**.

If the entire textual content of a document is considered, a **full-text search** is performed. However, if only the metadata or a specific part of the textual content, *i.e.* the author or title of a document, is considered, a **metadata search** is performed.

When performing a full-text search, several documents may match the search query, forming the results. However, the user is probably only interested in the most relevant results. In this case, the results must be ordered by their relevance to the query. This is called **ranking**. Ranking is the critical difference between information retrieval and retrieving results from a database. Compared to information retrieval, database searching only returns the objects that match the query in an arbitrary order, without any ranking. Figure 2.1 below shows the components of information retrieval and their interactions to build a system.

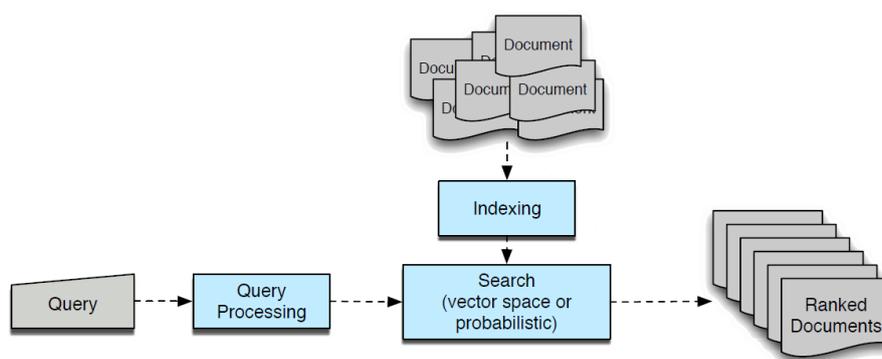


Figure 2.1: An overview of how information retrieval systems work. Image source: [12]

Full-text search can be further split into two categories:

- The first category is a **lexical search**, which stems from the word "lexicon," referring to a dictionary or the vocabulary of a specific language. A lexical search matches the words from a given dictionary without considering the meaning or context behind those words. Lexical search has further subcategories, such as **keyword search** that only matches the exact words in the query, **fuzzy search** and **wildcard search** also allow for typographical errors and word variations and lastly **proximity search** that instead of merely requiring the search terms to be present anywhere in the document, emphasizes the proximity of the terms to each other.

- **Semantic Search** carries out a search in a way that a lexical search can not. Using **text embeddings** generated by a neural network, the retrieval system can understand the context and meaning of the text. This allows for the search to be much more flexible *e.g.* a search for “canine” will produce results that do not contain the word “canine” but can include the word “dog”, as the two words have the same meaning.

To efficiently retrieve information, **indexes** are essential to an information retrieval system. An index is a data structure optimized for efficiently finding the relevant search results for a given query. These indexes are especially important when dealing with a large number of documents. With a small number of documents, it may not be necessary to use an index at all. Instead, the informational retrieval systems can directly scan the contents of the documents with each query. This strategy is called **serial scanning**.

The trading platform we develop in this thesis allows for the exploration of three different categories. Firstly, the companies themselves. Secondly, a company’s products, that do not necessarily have to be physical products but can also be services such as the provision of consulting. And lastly, a company’s needs, which is where a company can describe what it is are looking for *e.g.* a business partner, an investor or a specific product that fulfils certain criteria. We abstract the last two categories, products and needs, into listings for simplicity. Our platform describes companies and listings by a name and a description. We store additional details like location, industry, and contact information, but these are not used in searches. However, they could, in the future, be considered as filters for the search.

Therefore, in the context of our work, the task of exploring and searching for companies and listings can be defined as a full-text search. The query is the user’s intent, and the filters are the constraints the user wants to apply to the search results.

3 State of the Art

This section explores and discusses existing B2B trading platforms, text embeddings and their wide usage in state-of-the-art solutions. Lastly, the current state of information retrieval research and applications is reviewed. Understanding these cutting-edge advancements is crucial for developing an effective B2B trading platform that can meet the evolving needs of businesses in today's digital landscape.

3.1 B2B Trading Platforms

We analyse the search features of five of the largest B2B trading platforms: alibaba.com,⁷ crunchbase.com,⁸ clutch.co,⁹ upcity.com¹⁰ and wlv.de.¹¹

The aim of this analysis is to find answers to research question **RQ1.1** of which features may be relevant and beneficial to searching by comparing the current state-of-the-art B2B trading platforms. We also aim to compare these B2B trading platforms and their search features for our research question **RQ1.2** and to identify potential feature gaps within these B2B trading platforms that our platform can bridge in the future.

To analyse and compare the five B2B trading platforms, we visited each website manually and verified whether these platforms provided certain search features. We define four search features which we want to assess by means of a question for each. After defining each search feature, we summarize our findings and comparisons between the B2B trading platforms in Table 3.1 below.

Contextual Search

“Is it possible to search using detailed descriptions and natural language instead of just keywords?”

The contextual search feature is helpful for users who have an idea of what they want to search for and can describe their needs but do not know the relevant keywords. This forces the information retrieval system to understand the context of the search query rather than focus on keywords. We verified this feature by searching on the platforms with a description of a product instead of the product name itself. We deemed the feature unavailable if the search did not return the described product or something similar to what we know exists on the platform. An example of such a search is illustrated in Figure 3.1 below.

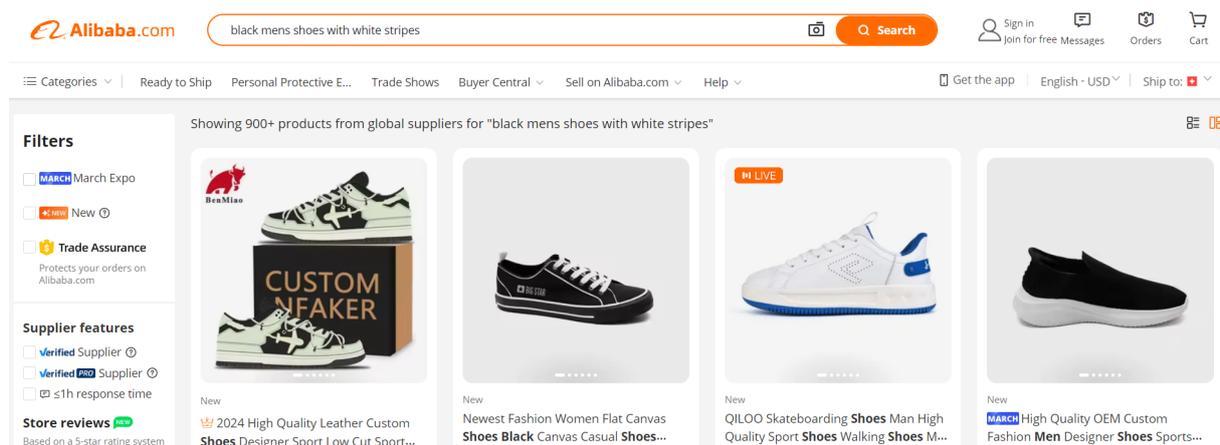


Figure 3.1: Example of the contextual search check on <https://www.alibaba.com/>, accessed: 19.03.2024

⁷<https://www.alibaba.com/>

⁸<https://www.crunchbase.com/>

⁹<https://clutch.co/>

¹⁰<https://upcity.com/>

¹¹<https://www.wlv.de/en/>

Result Type

“What type of search results are available to the user?”

Depending on the use case, users may prefer to search directly for products to purchase before contacting a company. In other cases, they may want to search directly for companies with which to partner and collaborate. Additionally, searching more granularly allows users to compare different results and potentially save time and money. We verify this feature by observing the possible different result types. We differentiate between the types “Product”, which means that the product is the prominent part of the search and “Company”, which means that the searches resulted in companies or other means of contact. In Figure 3.2 below, we show an example.

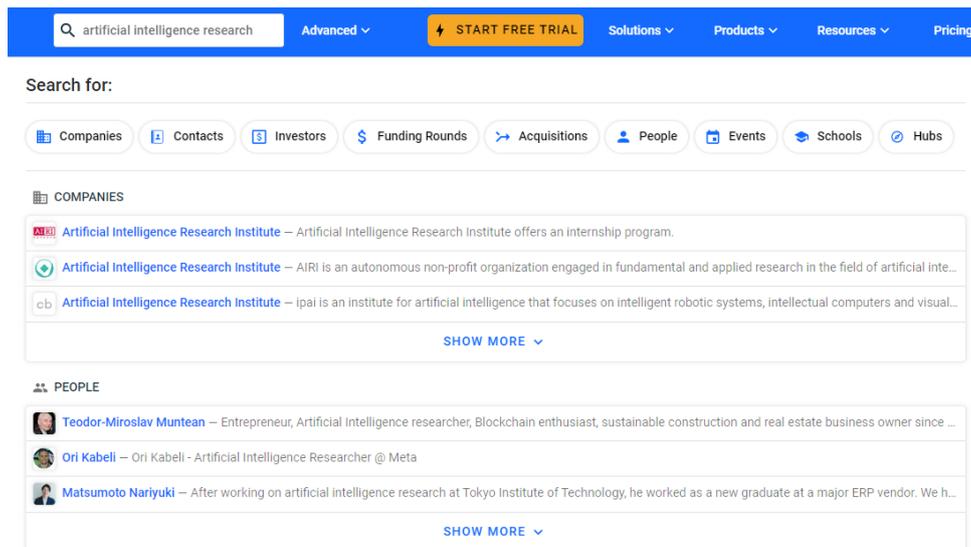


Figure 3.2: Example of a search on <https://www.crunchbase.com/> with the results of type “company”, accessed: 19.03.2024

Search Suggestions

“Does the platform provide search suggestions and autocomplete in its search?”

Search suggestions and autocomplete can not only speed up the searching process but also help users identify the most appropriate search terms and provide insight into whether the results are available or not. We verified this feature using partial searches or searches with typos and observed how the search bar reacted. We deemed this search feature unavailable if the search did not make any recommendations or attempts to complete or correct the search. Figure 3.3 below, illustrates an example of search suggestions.

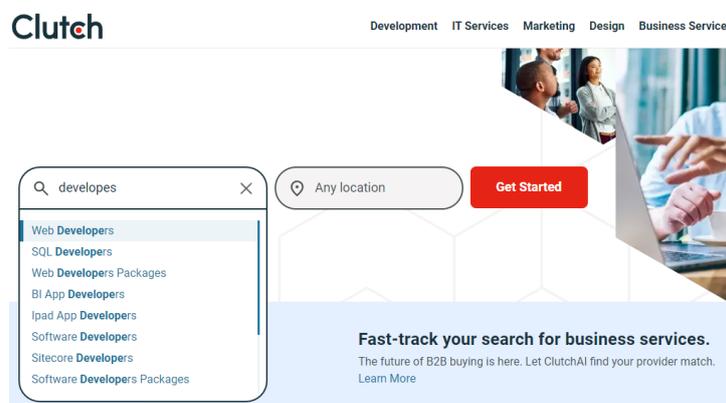


Figure 3.3: Example of search suggestions and auto-completion on <https://www.clutch.co/>, accessed: 19.03.2024

Multilingual

“Is the platform available in multiple languages?”

For international trading, translations of the user interface can support users who are not fluent in the website’s primary language. This search feature can dramatically improve usability and make a platform more approachable to new users. We verified this by searching for a button or option that allowed us to change the website’s language. Figure 3.4 below, shows such an option.



Figure 3.4: Example of an option to change the language on <https://www.wlw.de/>, accessed: 19.03.2024

Company	Contextual Search	Result Type	Search Suggestions	Multilingual
alibaba.com	X	PRODUCT	✓	✓
clutch.co	X	COMPANY	✓	X
crunchbase.com	X	COMPANY	X	✓
upcity.com	X	COMPANY	X	X
wlw.de	X	COMPANY	✓	✓

Table 3.1: Feature comparison of five of the current B2B trading platforms.

Our comparison of the five B2B trading platforms, as shown in Table 3.1 above, reveals some limitations in the current state-of-the-art trading platforms. Contextual search, a developing technology, is absent across all platforms. This is likely due to these platforms having been established several years ago and not updating their searches. Furthermore, most platforms prioritize companies as search results, with the exception of alibaba.com, which focuses on products. None of the five platforms offered both types of search results. Search suggestions, automatic search completion and multilingual support remain limited, with only three out of five platforms offering these features. These findings highlight the gap within these B2B trading platforms.

We also compare the current state-of-the-art B2B trading platforms to other search tools with the aim of identifying additional search features that could assist users when searching and thus answer research question **RQ1.1** [13][14]. We find that if B2B trading platforms could bridge these feature gaps and support other state-of-the-art search features, such as the below, they could dramatically improve their user experience and search accuracy:

- Highlight keywords or suggest corrections for typographical errors in their searches *e.g.* Google.¹²
- Maintain a search history such *e.g.* Youtube.¹³
- Offer filter suggestions based on query intent *e.g.* Ricardo.¹⁴
- Prioritize best-selling and relevant products *e.g.* Amazon.¹⁵

¹²<https://www.google.com/>

¹³<https://www.youtube.com/>

¹⁴<https://www.ricardo.ch/>

¹⁵<https://www.amazon.com/>

3.2 Text Embeddings

Text embeddings are essential to capturing the meaning of a word or sentence and can elevate searching to the next level. Text embeddings are vectors of a specific dimension generated by embedding models. These embedding models, for example Word2vec [15], are neural networks that have been pre-trained on a massive corpus of text and are able to capture the meaning of a word within a language. This capability allows the models to understand the nuances of human language, such as synonyms and subtle differences in meaning depending on the surrounding words.

These pre-trained models capture the meaning of a word by placing the generated text embeddings close within a defined vector space. For example, the embeddings for “engine” and “motor” would likely be closer than “engine” and “dog” due to their closer relation in meaning. This characteristic of proximity also results in the ability to use these vectors arithmetically. By performing arithmetic operations between text embeddings such as addition or subtraction, semantic relationships can be explored and words with similar meanings found, even if they have not been explicitly mentioned together before. This is a powerful tool for tasks such as finding synonyms and antonyms.

More advanced models, which are also the current state-of-the-art for text embeddings, known as sentence transformers [16], leverage the power of pre-trained models like BERT (Bidirectional Encoder Representations from Transformers) [17] to generate text embeddings over entire sentences or paragraphs not just single words, as shown below in Figure 3.5. These text embeddings are a fundamental building block of large language models [10] like ChatGPT¹⁶ and allow them to understand the relationships between words and capture not only the meaning of individual words but also the relationships and context within a text. This understanding allows them to perform complex tasks such as text generation, translation [18], and question-answering [19].

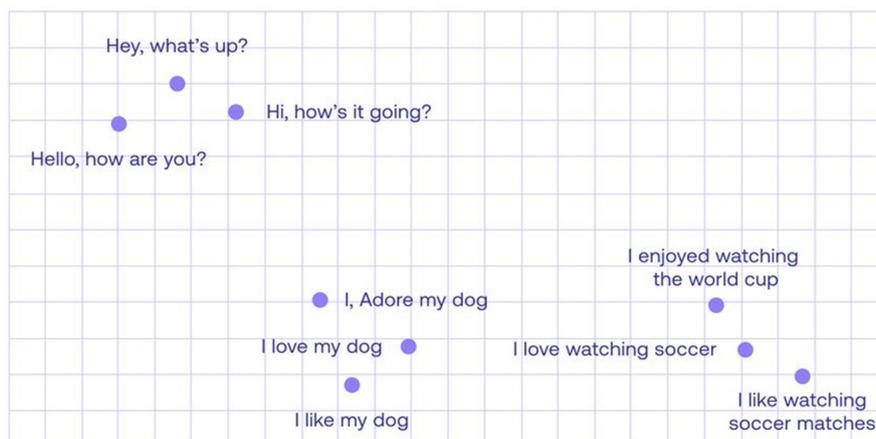


Figure 3.5: Visualization of text embeddings from a sentence transformer in a vector space.
Image source: [20]

¹⁶<https://chat.openai.com/>

By capturing the meaning and relationships within sentences, sentence transformers unlock a new level of text understanding and enable a wide range of powerful applications such as the following:

- **Semantic Search:** Rather than just searching for keywords, a semantic search [9] allows for a more reliable and widespread search due to an understanding of the relationships between words and their context in a sentence. This allows for the underlying meaning of user queries to be represented and improves retrieval of relevant results, making semantic search essential for our platform. Figure 3.6 below, shows how text embeddings are used to build a semantic search. Recent advancements with models like Retrieval-Augmented Generation (RAG) [21] further enhance semantic search by using text embeddings to identify relevant passages within a document from a private knowledge base and then leveraging a large language model (LLM) to generate a fact-based response without being prone to inventing information, also referred to as hallucinating [22] and having to perform expensive fine-tuning of LLMs.
- **Text Similarity:** The generated text embeddings of a sentence transformer can effectively be used to measure the similarity between texts. This capability is invaluable for tasks such as plagiarism detection [23], where the goal is to identify content that is too similar to existing sources *i.e.* by comparing the embeddings of a document to a database of known sources, potential instances of plagiarism can be identified. However, the significance of text embeddings extends far beyond plagiarism. Their ability to quantify semantic similarity unlocks a range of applications, including recommender systems [24] that suggest similar items to users based on their past interactions.
- **Multi-Modal Learning:** Embeddings bridge the gap between text and other data types such as images or videos by placing their embeddings in a common vector space. This opens doors to exciting applications in multi-modal learning, where models can process and understand different modalities together; for instance, image captioning [25], where a model generates a textual description of an image based on its content. Similarly, text embeddings can be used in text-to-image [26] or text-to-video generation [27], where textual descriptions are transformed into corresponding visual representations.

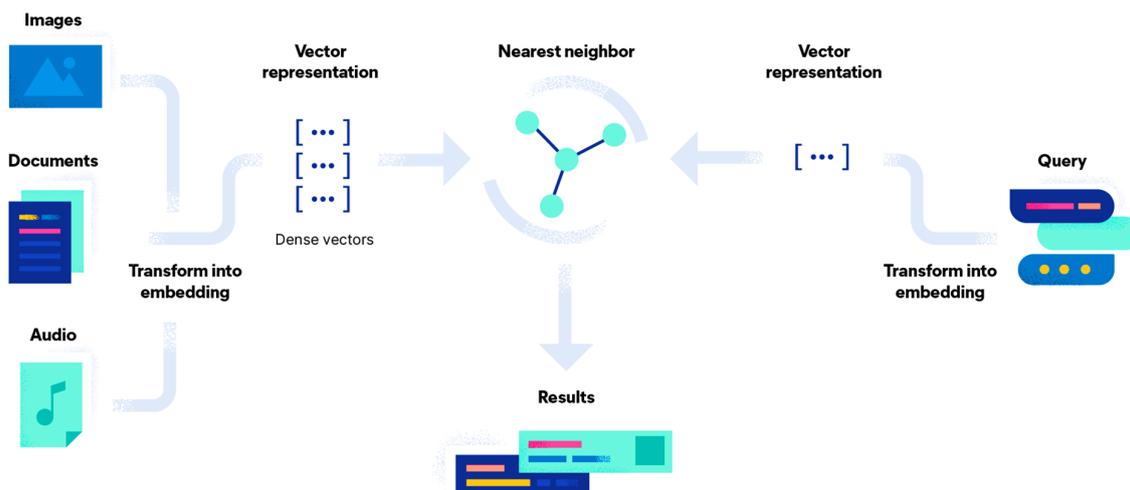


Figure 3.6: Visualization of how semantic search works to find relevant results. Image source: [28]

3.3 Information Retrieval

The primary objective of information retrieval (IR) is to locate and retrieve information relevant to a user's query. Apart from our use case of finding companies and listings, many other use cases can match this task, most of which we interact with daily. For example, current state-of-the-art information retrieval systems are the backbone of search engines [29], the e-commerce landscape [30], and recommendation systems [31], as both IR and recommendations rely on breaking down text into a representation that allows for comparison. Even for researchers to write about state-of-the-art work, as here, we use information retrieval systems to find relevant studies for our area of interest such as Semantic Scholar.¹⁷

Early information retrieval systems focused on lexical searches, such as keyword searching, where queries and documents were compared based on shared terms. This approach improved with techniques like lemmatization, which reduces words to their base or root form - lexemes - and allows for different word variations such as "running" and "run" to be treated equally. However, these methods struggle with synonyms, polysemy ("bat" as animal or baseball equipment), and context.

Google revolutionized searching the internet with graph-based approaches. Their PageRank algorithm [32], for instance, ranks webpages based on the quality and quantity of incoming links. This innovation was further enhanced by Google's Knowledge Graph [33], a massive database of entities and their relationships. When processing a search query, Google can leverage the Knowledge Graph to understand the context of the search query and identify related entities, leading to more relevant results.

The state-of-the-art in information retrieval has shifted towards semantic search [34]. This technique allows systems to move beyond simple keyword searching by capturing the meaning behind the words. This enables information retrieval systems to handle synonyms, polysemy, and context, overcoming the limitations of traditional lexical approaches. Even the advanced lexical searches using ranking metrics like TF-IDF (Term Frequency-Inverse Document Frequency) [35] or BM25 [36] struggle to compete with the power of semantic search. TF-IDF, as shown below in Figure 3.7, assigns weights to terms based on how often they appear in a document (term frequency) and how rare they are in the document corpus (inverse document frequency).

$$\begin{aligned} \text{tfidf}(q, d, D) &= \text{tf}(q, d) \cdot \text{idf}(q, D) \\ \text{tf}(q, d) &= \frac{f(q, d)}{|d|} \\ \text{idf}(q, D) &= \log \frac{1 + |D|}{1 + n(q, D)} \end{aligned}$$

Figure 3.7: Where q is the query term, d is the document, D is all documents in the corpus, $f(q, d)$ returns the number of times that the term q occurs in document d and $n(q, D)$ returns the number of documents that the term q occurs in.

¹⁷<https://www.semanticscholar.org/>

On the other hand, BM25 (BM standing for “best matching”), as shown below in Figure 3.8, enhances the TF-IDF weight by considering the document length and reducing the importance of terms that frequently appear in very long documents.

$$\text{bm25}(q, d, D) = \text{idf}(q, D) \frac{f(q, d) \cdot (k_1 + 1)}{f(q, d) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{\text{avgdl}})}$$

$$\text{idf}(q, D) = \log \frac{1 + |D|}{1 + n(q, D)}$$

Figure 3.8: Where q is the query term, d is the document, D is all documents in the corpus, avgdl the average document length in terms, $f(q, d)$ returns the number of times that the term q occurs in document d and $n(q, D)$ returns the number of documents that the term q occurs in. b and k_1 are free parameters, usually chosen as $b = 0.75$ and $k_1 \in [1.2, 2.0]$.

4 Our Approaches to Searching

Different information retrieval systems (IRS) have different capabilities. Therefore, in this section, we discuss why we chose these three information retrieval systems as our three approaches to searching. We then compare these three systems on a technical level to get an incline as to which one of the three would be best suited for our B2B trading platform. After analyzing the current state-of-the-art information retrieval systems, we decide to evaluate and compare the following three search approaches in our work:

- PostgreSQL Full Text Search (FTS) as our baseline.
- Semantic Search with pgvector for contextual and semantical understanding of search queries.
- Elasticsearch, the current industry standard.

In Table 4.1 below, we provide an overview of the three search approaches, where we analyze their search capabilities and assess their suitability for the various technical requirements of our B2B trading platform.

Feature	PostgreSQL FTS	Semantic Search	ES (FREE)	ES (PAID)
Keyword Search	✓	✓	✓	✓
Proximity Search	X	✓	✓	✓
Fuzzy Search	X	✓	✓	✓
Semantic Search	X	✓	X	✓
Advanced Ranking	X	✓	✓	✓
Easy Setup	✓	LIMITED	X	X
Scalability (Large Datasets)	LIMITED	LIMITED	✓	✓

Table 4.1: Comparison of our three selected information retrieval systems. ES is the abbreviation for Elasticsearch.

The choice of approach depends on the user’s specific needs. If a basic keyword search is sufficient, PostgreSQL FTS is very reliable. For more advanced features such as proximity search, fuzzy search, and semantic search, pgvector or Elasticsearch are better options. The tradeoff between these systems is the ease of setup and maintainability versus the option to scale horizontally for large datasets. Below, we further discuss the internal workings of these three search approaches.

4.1 PostgreSQL Full Text Search

PostgreSQL Full Text Search (FTS) is often chosen due to its ease of use and straightforward setup, making it ideal for adding basic search functionality to applications. It is especially popular as it is well-documented [37] and built into all PostgreSQL databases. However, PostgreSQL FTS has its limitations. It struggles with handling typographical errors and variations in word forms, *e.g.* “machinery” vs “machine”. As the length of text increases, the accuracy of search results can decrease. This is because individual keywords (lexemes) become less meaningful in isolation as the length of text grows. Additionally, PostgreSQL FTS relies on indexes that become very expensive to compute with large datasets, limiting its scalability capabilities [38].

4.2 Semantic Search with pgvector

Pgvector is an extension for PostgreSQL that enables the storage of high-dimensional vectors and facilitates semantic search. It relies on an embedding model, such as a sentence transformer, to generate the document and query text embeddings. These embeddings are high-dimensional vectors representing the underlying information. While pgvector enables the storage of these vectors, it can not do so for an arbitrary vector dimension. If vectors have more than 2'000 dimensions, then the storage is no longer possible.¹⁸ For semantic search, pgvector adds three new operators for the corresponding distance functions used to calculate the similarity of the query and the document vectors. The pgvector extension supports the following distance functions:

- Cosine distance
- Euclidean (L2) distance
- Inner or dot product

4.3 Elasticsearch

Elasticsearch is considered the industry standard for document retrieval and full-text search as it is highly scalable and offers advanced ranking algorithms. However, it is important to be aware of its limitations. The free version has restricted features, such as not allowing semantic search. Setting up Elasticsearch can also be complex, as it requires additional dependencies like Kibana¹⁹ for the user interface and Logstash²⁰ or pgsync²¹ to ingest and synchronize data between databases. Its use of “shards”, which split up the search indexes and distribute them across multiple servers, illustrated in Figure 4.1 below, allows for excellent scalability with large datasets. While it provides promising search capabilities and allows for a platform to scale, it requires significant configuration. Elasticsearch can also be very expensive to run and maintain, especially when utilizing the advanced features offered in the paid version.

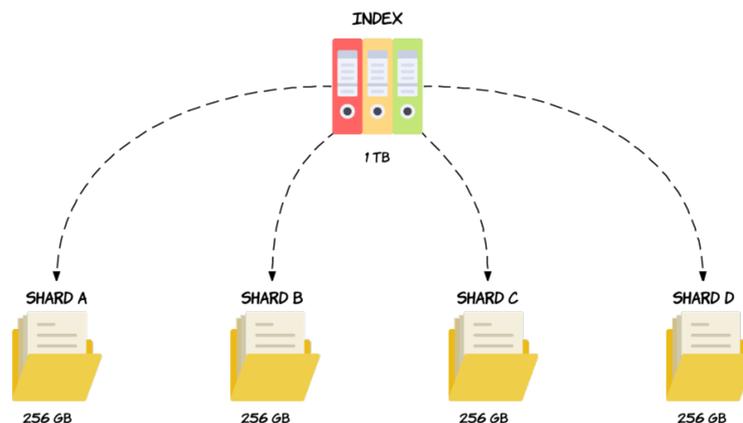


Figure 4.1: Visualization of how Elasticsearch creates an index and splits it into shards. Image source: [39]

¹⁸<https://github.com/pgvector/pgvector?tab=readme-ov-file#what-if-i-want-to-index-vectors-with-more-than-2000-dimensions>

¹⁹<https://www.elastic.co/kibana>

²⁰<https://www.elastic.co/logstash>

²¹<https://pgsync.com/>

5 Our Dataset

In this section, we discuss the data requirements for our B2B trading platform and how we acquired a suitable dataset.

It is not possible to implement and evaluate our three approaches to searching until the platform is populated with data. We define three key requirements that must be fulfilled by the data for our platform:

1. The data must contain natural language. For simplicity, this natural language should also only be in English. If we wanted to support multiple languages, the search functionality would become much more complex. However, in the future development of this platform, this might be a factor that has to be considered, as the goal of our platform is to support international B2B trading.
2. The data should match the structure of our platform. At the minimum, our platform has companies with names and descriptions. These companies can then offer their products or services. A company can also describe their needs or desires, which other platform users may be able to fulfil. For simplicity, we refer to both these last parts of a company as “listings”, which also require a name and a description. We also use the listing abstraction in our code.
3. The data should cover a wide variety of business sectors as the platform will not be limited to only one specific business sector.

Due to our focus being mainly on the search functionality, we can not use placeholder data such as “lorem ipsum”.²² If placeholder data were used, the search functionality would not make a lot of sense because the search queries we use are natural language which “lorem ipsum” is not. For a demonstration of the platform to potential investors or partners, placeholder data is also not very attractive and makes it hard to understand the platform’s objective.

After defining the above requirements for our data, we identified three possible methods to source our data. Each of these three methods is explained below.

5.1 Existing Datasets

There are many datasets which are widely available, Huggingface²³ & Kaggle²⁴ being the two largest providers. Using a pre-existing dataset is the most common approach for research as it significantly reduces the time and resources required for data collection, preparation and labelling whilst also allowing for comparison with other research using the same dataset. However, these datasets can also have certain biases. For example, one of the most commonly used datasets for image classification, ImageNet,²⁵ has been shown to have a bias for textures in the image [40]. These biases can impact a solution’s performance and introduce potentially harmful stereotypes if not carefully addressed [41]. One limitation of the approach of using an existing dataset is the availability of suitable datasets. In our case, we could not find a dataset that met the aforementioned data requirements, rendering this approach infeasible.

5.2 Scraping Data from other Platforms

Data scraping involves extracting information from websites, either manually or automatically. To automatically extract the embedded data, the website’s source code must be reverse-engineered. However, this process can be quite difficult and tedious due to the minimization of source code to save bandwidth while serving the website. Another hurdle is finding a website or multiple websites that provide the

²²<https://www.lipsum.com/>

²³<https://huggingface.co/>

²⁴<https://www.kaggle.com/>

²⁵<https://www.image-net.org/>

relevant information. Due to our defined data structure of companies, products and needs, we could not find suitable websites providing matching data.

Another aspect to consider is legality when it comes to scraping. While data mining is allowed for information generation, sharing the contents can result in legal issues [42]. As we will be using this data while presenting our work, we have decided not to use this method to avoid any potential legal issues.

5.3 Creating our own Dataset

Creating a dataset from scratch is another approach that usually costs a lot of time and money and is, therefore, not usually a viable option. However, since the groundbreaking paper on transformer models [8], the landscape of natural language processing (NLP) has undergone substantial changes. With these powerful transformer models, natural language generation has become remarkably efficient, with applications like ChatGPT²⁶ or Github Copilot²⁷ becoming part of our daily life for many. These large language models (LLMs) [10] excel in capturing intricate patterns and dependencies within natural language, enabling them to produce coherent and contextually relevant text. Therefore, in order to populate our platform with data, we decided to use the OpenAI's GPT-4 LLM.²⁸

Initially, we created a list of possible business sectors that our platform should cover and added a description for each sector, as can be seen in Listing 1 below.

```
{
  "name": "Information Technology (IT) and Software Development",
  "description": "This sector focuses on the development,
                 implementation, and management of
                 computer systems, software, networks,
                 and information services. It includes
                 industries like software development,
                 IT consulting, computer hardware
                 manufacturing, telecommunications, and
                 data analysis."
}
```

Listing 1: Example of a business sector and its description.

In order to create a rich but still diverse dataset, we subsequently chose to restrict ourselves to five of these business sectors. We based our choice on those sectors that could benefit from such a platform and on sectors that could contribute the most to the UNESCO and Sustainable Development Goals.²⁹ The five business sectors we ultimately defined are as follows:

- Energy and Utilities
- Financial Services
- Healthcare and Pharmaceuticals
- Information Technology (IT) and Software Development
- Manufacturing

Based on these five business sectors of interest, we built up a pipeline that generates companies and matching listings. This pipeline was written in Python, the preferred language for working with machine learning or deep learning algorithms. We used the instructor package³⁰ which is a wrapper around

²⁶<https://chat.openai.com/>

²⁷<https://github.com/features/copilot>

²⁸<https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo>

²⁹<https://en.unesco.org/sustainabledevelopmentgoals>

³⁰<https://github.com/jxnl/instructor>

OpenAI's API package.³¹ The Instructor allows us to constrain the large language model's generated text to a structured output. The structured output is defined by Pydantic ³² models, which are type-checked Python models. Pydantic also allows for the addition of validators to a data model's field, which are checked during object construction. The Instructor forces the LLM to output JSON. An attempt is then made to parse the JSON to the provided Pydantic model. If the parsing fails due to limitations or hallucinations of the LLM, it is possible to set a number of times the Instructor should retry to generate a valid output, as can be seen in Listing 2 below. By having structured output from the LLM, we can ensure that our data is consistent and can be easily iterated and worked on.

```
from pydantic import BaseModel

class CompanyResponse(BaseModel):
    name: str
    description: str
    ...
company = client.chat.completions.create(
    model="gpt-4",
    temperature=0.8, # 0 = deterministic, 1 = random
    response_model=CompanyResponse,
    # number of times to retry in case
    # the response doesn't match the response_model
    max_retries=10,
    messages=[...]
)
current_company = Company(
    name=company.name,
    description=company.description,
    sector=current_sector
)
...
```

Listing 2: Generating structured output with the Instructor package and parsing it to a valid Pydantic model.

³¹<https://github.com/openai/openai-python>

³²<https://github.com/pydantic/pydantic>

5.3.1 Prompting Chain

With the structured outputs generated in the method previously mentioned, we built a chain of prompts that iteratively builds up our desired data structure, as can be seen below in Figure 5.1. Within the chain of prompts, there are multiple different prompts. We experimented with many variations of prompts, such as generating an entire company or listing at once, but this was very challenging for the LLM. For each prompt in the chain, we also experimented with different formulations and common patterns that are suggested by OpenAI³³ or the researching community has found [43]. Empirically, we concluded that chain-of-thought prompts [44] that make the model think step by step and gradually build up the data structure provide the best results. All prompts that we used to create our dataset can be found in Appendix A or the corresponding Github repository.³⁴

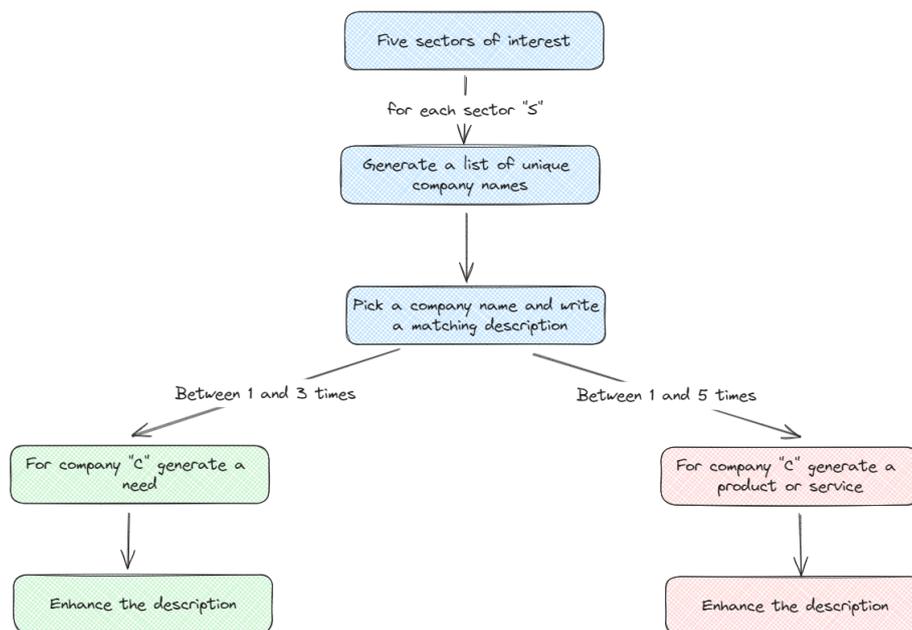


Figure 5.1: A simplified diagram of the prompting and generation process for a company (blue), the products and services it offers (red) and the needs and desires it seeks (green).

³³<https://platform.openai.com/docs/guides/prompt-engineering/strategy-write-clear-instructions>

³⁴<https://github.com/LuciferUchiha/SMEConnectGPT>

5.3.2 Data Analysis

After generating our dataset, we qualitatively analysed the resulting data. Despite adjustments to the prompts, the quality of the data for the needs of companies consistently fell short compared to that for the companies themselves and their products. In some instances, the model even replicated the company descriptions provided as context. Due to this quality disparity, we excluded company needs from further evaluation, focusing on companies and products only.

We also analyzed the text embeddings of our synthetically created data. As illustrated in Figure 5.2 below, the business sectors significantly influenced the descriptions of the generated companies, products and needs. This is evident by the tight clustering of each sector when reducing the vector's dimensionality from 354 to 2 using UMAP [45] and its Python package.³⁵ The smaller density of the company need embeddings can be interpreted as the drop in quality.

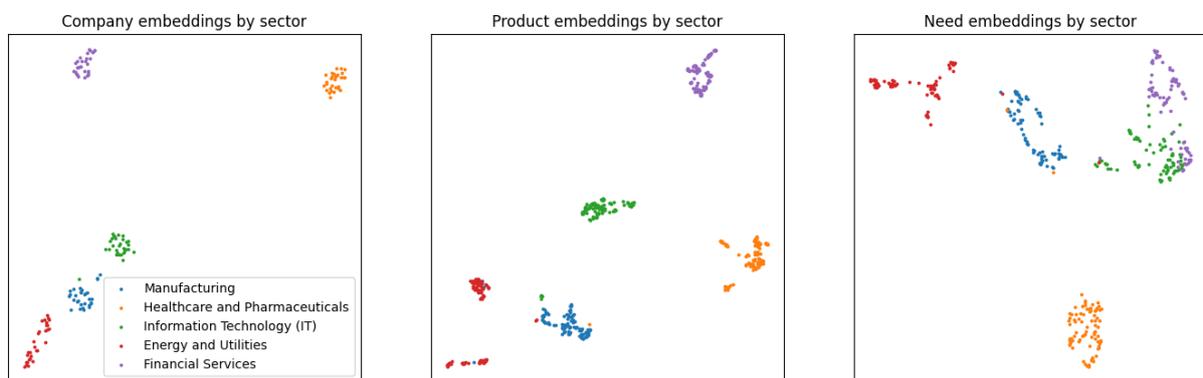


Figure 5.2: Visualization of the text embeddings after being reduced to two dimensions.

5.3.3 Data Enhancements

Later in the development process, we also decide to enhance our data using the Faker³⁶ package. The Faker package provides fake but realistic data for specific fields such as an email address, country, city *etc.* This package was used to add additional fields to the companies and listings such as location and contact information and to emulate the real data the platform could be populated with in the future.

³⁵<https://github.com/lmcinnes/umap>

³⁶<https://fakerjs.dev/>

6 Conceptual Solution

This section outlines the platform’s conceptual design choices, laying the groundwork for its technical details and implementation. We will present the structure of our data models, explain why we use certain frameworks and technologies and lastly discuss the overall architecture of the platform. We also showcase our user interface wireframes.

6.1 Data Models

Our platform comprises of the four data models shown in Figure 6.1 below: companies, listings, users, and sectors. These models have many interconnected relationships. For example, a single company can have multiple listings (One-to-Many) and be active in multiple sectors (Many-to-Many). Currently, the company model has the most fields, but this may change in the future, as certain tables can be separated and shared between multiple classes, such as contact information. The most crucial fields are the names and descriptions of both the companies and listings because these are relevant to the information retrieval systems to facilitate the platform’s search and exploration functionality. All data models apart from the user were populated with our previously created dataset, as described in section 5.

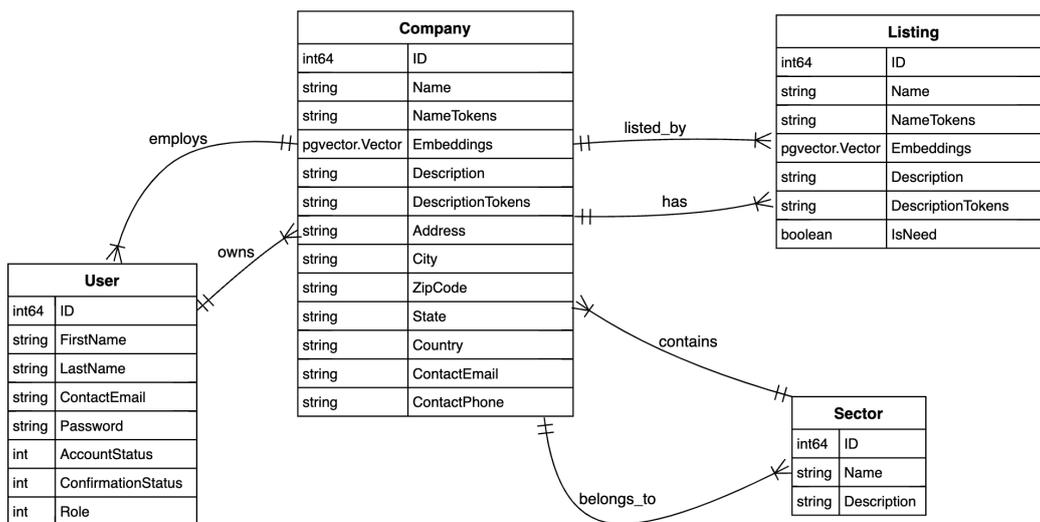


Figure 6.1: Our data models.

In the overall user scenario, a user can have products and/or services they want to offer on the platform and needs and desires they want fulfilled. To represent this separation of interest, we define *needs* (for the listings a company wants) and *products* (for the listings a company offers). We use the *is_need* field to represent this split in the data model. To handle these relationships, we use companies where a single user is the owner but other users can also join the company. We believe this will discourage account sharing and ensure a more secure platform while allowing users to pass on the companies to their successors.

6.2 Frameworks and Programming Languages

Choosing the right programming language for the business logic of the platform was crucial, as it can have a long-lasting effect on the construction, performance and maintainability of the platform. Especially because primarily, students will be developing the platform. We explore, the following four options:

- **Java with Spring Boot:**³⁷ Spring Boot can be a powerful framework for implementing APIs. However, the implementation can become very complex depending on the use case. Spring boot relies on many other heavy dependencies, which can slow down the application. Spring Boot also takes over a lot of implementation details for the developer, which can be very useful but can also lead to unexpected behaviour or workarounds, which, if left undocumented, can later lead to issues with different students working on the platform.
- **Rust:**³⁸ This language has gained popularity recently and offers impressive performance and memory management. However, its complex low-level syntax, paired with a high learning curve and ever-evolving library, was deemed too challenging for a student-oriented development environment.
- **Functional Programming Languages (Elixir,³⁹ Haskell,⁴⁰ OCaml⁴¹):** While these languages are known for their speed and concurrency, learning them can be challenging as they use a different approach to most programming languages. This unfamiliar approach, based on functional programming concepts, may require students to have some background knowledge they may not have.

Ultimately, **Go**⁴² emerged as the ideal choice due to its fulfilment of the following criteria:

- **Flexibility and Performance:** Go allows for polymorphic implementation, enabling a more structured project growth and adaptation while maintaining high performance thanks to its efficient compilation.
- **Learning Curve:** Compared to the other three options, Go presents a significantly flatter learning curve for students, which is crucial if there are frequent team changes.
- **Simple and Transparent Code:** Unlike the Spring Boot framework, which hides many implementation details, Go has a clear logic flow that prevents the hiding of implementation details.
- **Error Handling:** Go's "errors as values" idiom simplifies error management, making it easier for students to identify and handle potential issues within the code. This, together with Go's convention of "no fail silently," encourages error handling, which in turn promotes secure applications.
- **Concurrency and Scalability:** Go inherently supports concurrency, which is crucial for an API to handle multiple requests simultaneously. Go's built-in concurrency features, such as lightweight goroutines, also make it horizontally scalable.
- **Memory Management:** Go has an efficient Garbage Collector (GC), which frees developers from manually allocating and freeing memory that is often forgotten and leads to memory leaks and crashes. However, Go still allows developers to use pointers, allowing for more memory-efficient implementations.
- **Rising Popularity and Community Support:** Go's growing popularity, backed by Google as its creator, ensures an active community for students to seek guidance and support throughout the development process.

³⁷<https://spring.io/projects/spring-boot>

³⁸<https://www.rust-lang.org/>

³⁹<https://elixir-lang.org/>

⁴⁰<https://www.haskell.org/>

⁴¹<https://ocaml.org/>

⁴²<https://go.dev/>

With its well-rounded feature set, Go provides the perfect balance for our project, enabling both efficient development and a suitable learning experience.

For the user interface, we considered using backend-generated templates. However, this approach tightly couples the user interface with the business logic, making it unsuitable as we wanted the platform to remain flexible. We decided that to enable a highly responsive user experience during search and exploration of the platform, a framework allowing the development of a single-page application (SPA) would be best suited. This approach prioritizes client-side processing of user interactions, significantly enhancing perceived response times compared to server-centric architectures like templates. As the name suggests, a SPA simulates the multi-page experience users are familiar with by dynamically manipulating the initial page, enabling a seamless and interactive user experience.

While frameworks like Angular,⁴³ Svelte,⁴⁴ Solid,⁴⁵ and Vue⁴⁶ all build SPAs and offer compelling features, our lack of experience with these frameworks led us to choose React.⁴⁷ React's presence in courses at FHNW and our existing experience, coupled with its extensive library ecosystem, made it the ideal choice. We use React with TypeScript,⁴⁸ a superset of JavaScript that adds optional typing to improve code readability, maintainability and security.

Next.js⁴⁹ emerged as the clear favourite among React's available frameworks due to its documentation, abstractions for complex React features, rich library ecosystem, built-in caching, server-side rendering solution and developer tooling.

For data fetching, we use React Query.⁵⁰ This library provides the following benefits:

- **Manages loading states:** React Query exposes the "loading" state, enabling users to receive feedback on what is happening within the platform.
- **Supports optimistic updates:** It enables optimistic updates, which temporarily reflect changes on the client side before server confirmation, enhancing user experience.
- **Reduces useEffect usage:** React Query minimizes the need for the `useEffect` hook, a known source of bugs in data fetching logic. This helps to improve code maintainability and reduce potential errors and crashes.

Furthermore, we opted for Tailwind CSS,⁵¹ a utility-first styling framework. Tailwind CSS offers several advantages over other frameworks to style the user interface as follows:

- **Reduced bundle size:** By providing small, reusable utility classes instead of extensive CSS rules, Tailwind CSS minimizes the final CSS bundle size using a compiler that merges them at build time.
- **Direct component styling:** Tailwind CSS classes can be directly applied to the React components, eliminating the need for separate CSS files and making the code more readable and maintainable.

We ultimately decided to use the shadcn/ui⁵² component library for our project. Unlike other libraries, which only provide pre-built components, shadcn/ui offers access to the complete source code of the components, allowing for easy modification. Additionally, its use of Tailwind CSS perfectly aligns with our chosen styling solution. Our decision to use this library resulted in an efficient user interface development process, minimizing complex CSS work and achieving a visually appealing and consistent user interface.

⁴³<https://angular.io/>

⁴⁴<https://svelte.dev/>

⁴⁵<https://solidjs.com/>

⁴⁶<https://vuejs.org/>

⁴⁷<https://react.dev/>

⁴⁸<https://www.typescriptlang.org/>

⁴⁹<https://nextjs.org/>

⁵⁰<https://tanstack.com/query>

⁵¹<https://tailwindcss.com/>

⁵²<https://ui.shadcn.com/>

6.3 Monolith vs Microservice Architecture

Choosing between a monolithic and microservice architecture is a critical decision for this project. While both approaches have their merits, our decision hinges on fulfilling specific requirements. Our primary concern is ensuring a fast and efficient search function using our three search approaches and allowing the platform to scale. Additionally, the project will, in the future, involve multiple development teams, necessitating a structure that facilitates independent work. Finally, we aim for a high degree of agility, allowing for quick and easy code changes as needed. Considering these factors, a careful evaluation of both architectural styles is needed. In Figure 6.2 below, we show a simplified visualization of the two architectures.

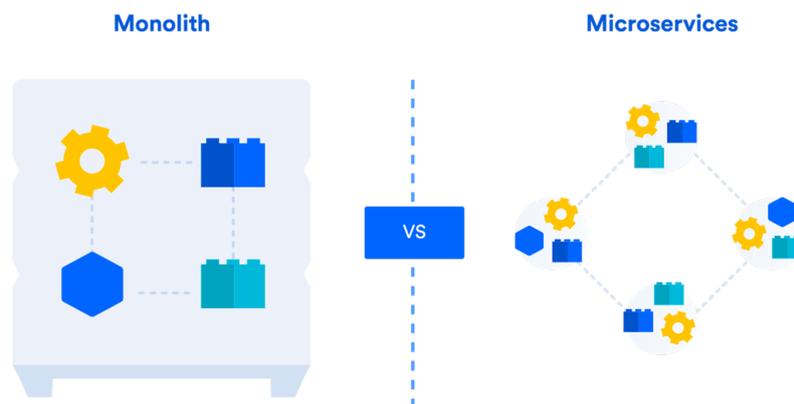


Figure 6.2: A visualization of the difference between a monolith and microservices. Image source: [46]

The most significant advantage of monoliths is the simplified setup and lower internal latency due to the fact that there is no additional serialization for transportation. This is because the code shares the same memory address space. The biggest weakness of monoliths is that they can only scale vertically, *e.g.* the only possibility to increase performance is to improve the server hardware or the logic within the platform.

On the other hand, microservices allow for horizontal scaling. Each service can be independently built, requiring only a defined communication contract. However, this separation introduces additional complexity. The communication layer itself also adds complexity and potential latency. Using integration testing, the inherent risk of coupling the different systems together can be minimized.

Below, we will provide further details as to the reasons behind our choice of the microservice architecture.

Flexibility: Over time, the platform will be developed and extended by many students with different areas of expertise and different levels of knowledge regarding programming languages, frameworks and architectures. When future developers take over this project, they first have to understand what has previously been developed before they can work on it. However, due to the microservice architecture, they would not need to understand the internal workings of existing services. When other students want to extend the platform, they can develop their own microservice in their preferred programming languages and only need to understand how to combine the different services. This results in microservices simplifying the integration of different programming languages compared to a monolithic architecture. Using, for example, Python for a new NLP service alongside existing languages could be complex in a monolithic system. However, the decoupling of the services allows them to be easily exchanged, so students can explore different technologies and solutions without any issues as long as the API stays consistent and the contracts are fulfilled.

Scalability: In the case of this platform, each service has a different set of computational requirements and impact on the platform. For example, the search functionality is expensive to compute and essential to the platform. Therefore, the platform could benefit from having multiple instances of the search services available. This can also be achieved dynamically by using auto-scaling, where instances are started and stopped as the platform load increases and decreases, as shown in Figure 6.3 below. This is only possible due to the composability of microservices. With a monolith, we would have to scale the whole system, as it is not possible to only scale certain parts of it. Automatic scaling can also help to cut costs and reduce energy consumption by only using resources when required.

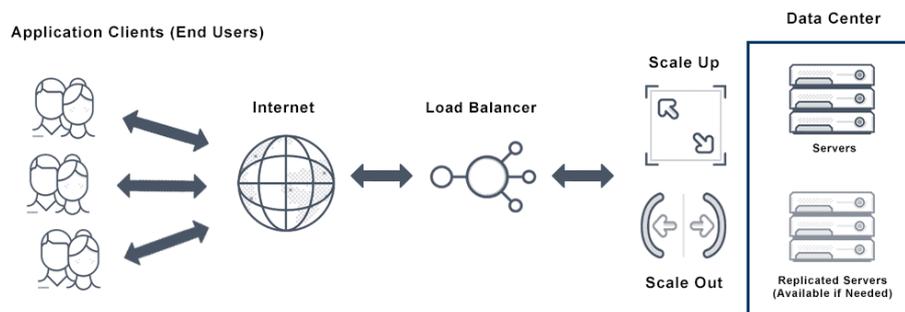


Figure 6.3: How auto-scaling works in a simplified way. Image source: [47]

Independence: The platform can be split into the following independent services that each fulfil a different part in our separation of concerns:

- **The User Management Service (UMS)** manages the users and provides the authentication layer.
- **The Company / Listing Management Service (CLMS)** manages the companies and the listings. It also hosts the implementation of the information retrieval systems.
- **The Embedding Service (ES)** generates the embeddings *i.e.* latent vectors for the semantic search.

Although splitting the CLMS into separate services for companies, listings, and sectors was an option, we chose to keep it as a single service. This is because the data models are closely interconnected. Separating them would create more complexity and latency and add possible sources of errors by introducing an additional communication layer.

Due to the similar natures of some services, a further grouping of our services can be established:

- **The controller layer** consists solely of the backend, which accepts the incoming traffic and acts as the controller to the outside for the microservices and then routes these requests to the next layer.
- **The application layer** is where the business logic of our platform is located. The layer receives the requests from the controller layer and processes them depending on the specific service *e.g.* UMS or CLMS.
- **The data layer** contains all services related to the data storage and processing, *e.g.* the Postgres database, the Elasticsearch (which also serves as a database) and the embedding service responsible for creating the vectors used for the semantic search. We use different parts of the data layer for each information retrieval system.

Figure 6.4 below shows an overview of the platform and these layers. The data flow of the platform is also visible and it also illustrates that neither the front- nor the backend has any direct connection to the database, increasing data protection.

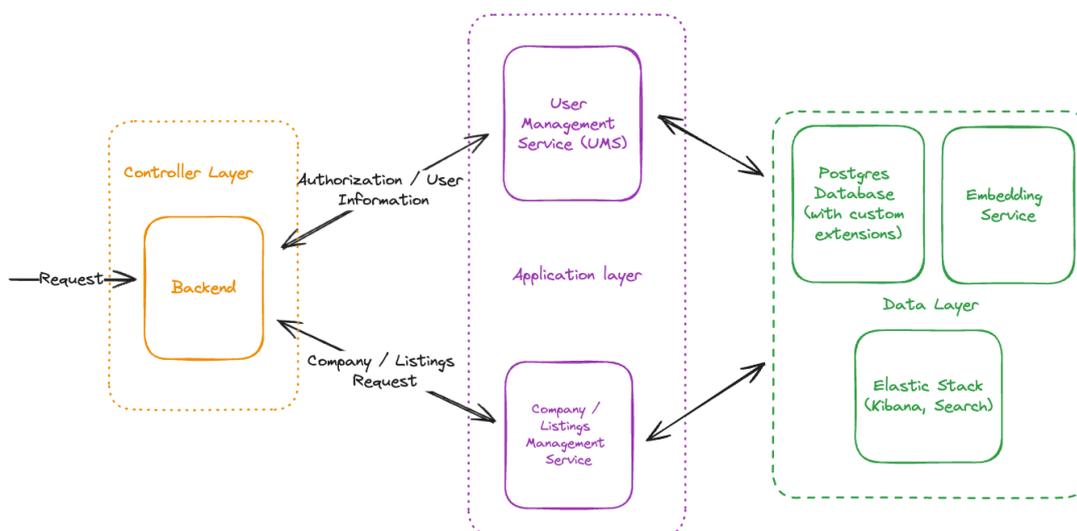


Figure 6.4: Layering of our microservices.

6.4 User Interface Design

While the ultimate goal is a polished user interface, for this initial development phase, we focus on creating wireframes for the design process. Wireframes are used because they allow the prioritization of functionality and information architecture over visual design, which in turn facilitates a rapid iterative design and development process. This approach also allows us to focus on the implementation of information retrieval systems' functionalities, as we believe that the user interface primarily serves as a visualization tool for user interaction with the underlying information retrieval systems.

Figure 6.5 below presents the wireframe for the search page. Our priorities for the search page were the following:

- **Targeted Filtering:** At the top of the search page, we provide the user with three main search filters for companies, products and needs. This allows for a more targeted search experience tailored to specific user goals.
- **Conversational Search:** We positioned the search bar at the bottom of the interface. This unconventional approach draws inspiration from chat applications like ChatGPT, where search interactions often resemble conversations. We believe this is a potential future direction for searching, where a more natural, question-and-answer style is likely to become the norm.
- **Streamlined Result Exploration:** We display the search on the left side of the screen, leaving most of the screen for a detailed view of a selected result. Within the detailed view, the user could, in the future, then browse, moving, for example, from a product to its company and back without affecting the search results on the left. By designing the search page this way, the need to open and maintain multiple tabs or navigate back and forth between web pages can be eliminated. This has the potential to drastically streamline the user experience.

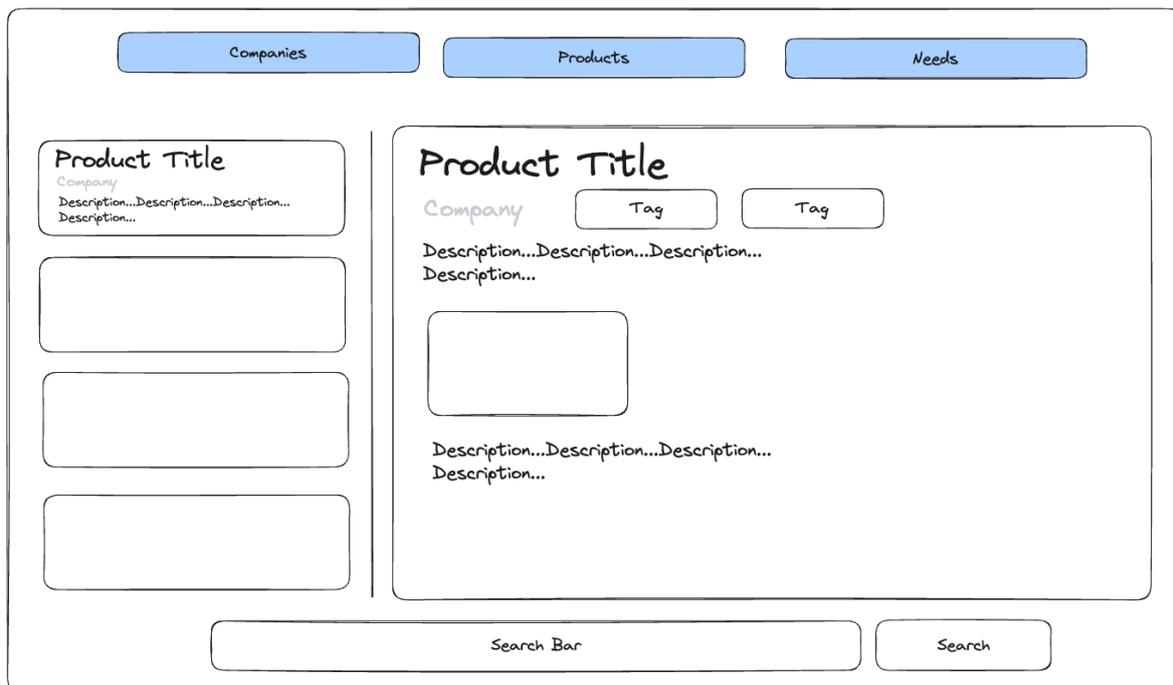


Figure 6.5: Wireframe for the search page.

In the interest of completeness, we also designed a potential onboarding process that details the steps from registering a user account to setting up a company profile. These wireframes were not implemented in the current phase due to their lower value and priority in comparison to the core search functionality. However, we include them in Appendix B for the benefit of future students. In our implementation, we provide a simple login page to add a layer of security for when the platform is deployed.

7 Solution

Having established the gap in the B2B trading market and explored our choices on how to build a conceptual bridge with our platform, we now focus on how we implemented our platform and integrated our three search approaches. This section delves into the key decisions made during development and the challenges encountered.

7.1 Frontend

Our frontend translates user actions into instructions for the business logic and presents the results to the user in a clear and understandable way. Below, we briefly illustrate and discuss the implementation of our platform's user interface.

7.1.1 Login Page

The aim of the login page is to add an additional layer of security by requiring a user to be registered on the platform. Only developers can create user accounts, preventing unauthorized access to the deployed platform by requiring a valid username and password. This layer became especially important when our second iteration of the embedding service utilized OpenAI's API to generate the text embeddings for the semantic search. This usage presented a cost concern, as each API request incurs a fee. Implementing an authentication layer allows us to control user access and limit unnecessary API calls.

For our authentication layer, we use the Auth.js framework.⁵³ Auth.js streamlines the authentication process by providing out-of-the-box features like user sessions, email verification, and password hashing. The framework also offers support for various third-party authentication providers, such as Google and Facebook, through OAuth adapters.⁵⁴ Using OAuth adapters, we could eliminate the need to create and remember separate usernames and passwords that are easily forgotten. We also increase security when introducing an OAuth adapter as most providers support two-factor authentication (2FA) and because OAuth reduces the risk of users' passwords being stolen. Even if a security breach occurs, attackers would not have access to the users' main account credentials. Using an OAuth provider also reduces the development and platform complexity with regard to account management, as the responsibility lies with the provider.

⁵³<https://authjs.dev/>

⁵⁴<https://oauth.net/2/>

7.1.2 Search Page

In Figure 7.1 below, our implementation of the search page is illustrated, matching our previously designed wireframe and ideas in subsection 6.4. The consistent design is clearly visible thanks to the shadcn/ui component library.

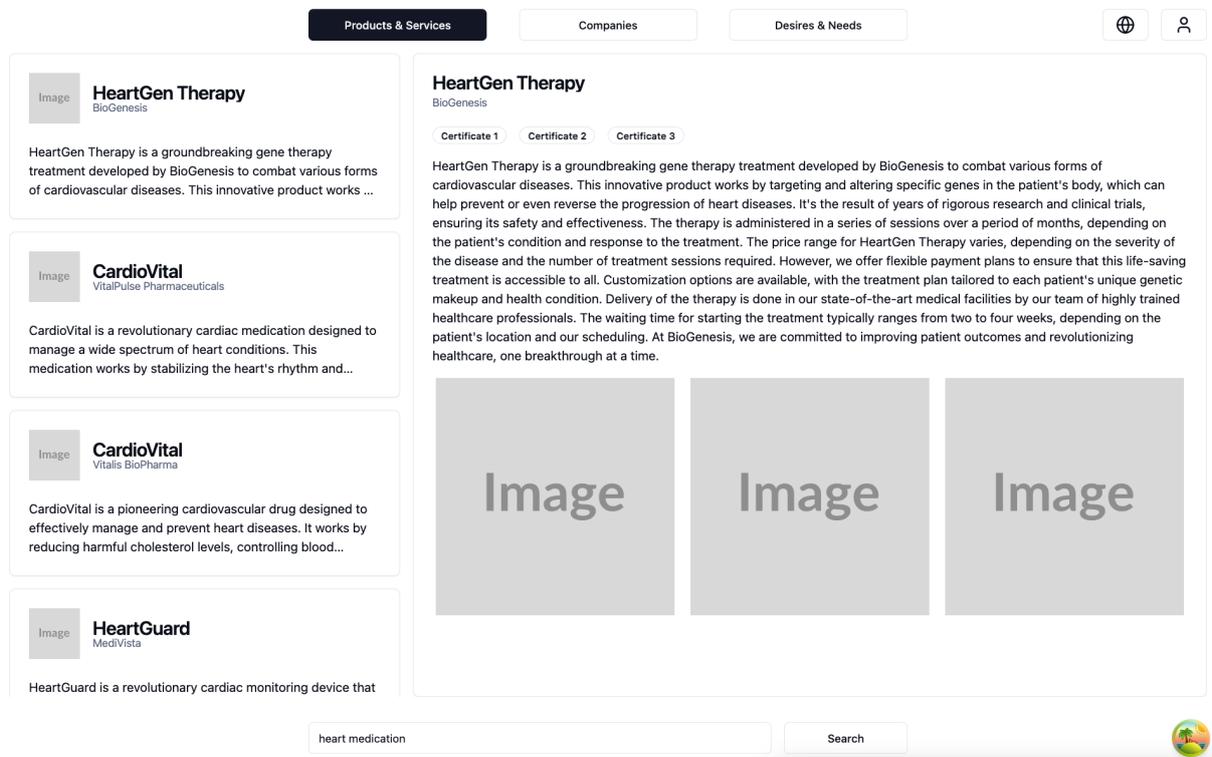


Figure 7.1: The user interface for the search page. At the top are the category filters; at the bottom, the search bar; on the left are the search results and in the center are the details of the selected search result. The globe in the top right corner is used to set the language.

7.1.3 Multi-Language Support

The platform is intended to be used internationally, and therefore, we lay the groundwork for multi-language support. To incorporate multi-language support, we use the `next-intl` framework.⁵⁵ The framework requires some initial configuration, such as setting the default language. However, after the setup, we can define for each supported language a dedicated JSON file, as shown in Listing 3 below. This JSON is then provided to `next-intl` by placing it in the `translations` folder. Currently, the platform supports English and German, but expansion is straightforward by adding a new JSON file for each additional language.

The user's browser preference sets the initial language. If nothing is set, then our defined default language, English, is used. The user can then override the language using the dropdown menu that is present in the top-right corner of the application, as shown in Figure 7.1 above. The selected language not only changes the text but also has an effect on the website's URL. For example, a user viewing the platform in English would see the URL `[base_url]/en/search`, while a German user would see `[base_url]/de/search`. Here, the `[base_url]` represents the main platform domain (e.g., "mycompany.com"), and `en` and `de` represent the language codes for English and German, respectively.

```
{
  "Login": {
    "title": "Sign in to your account",
    "usernamePlaceholder": "Username",
    "passwordPlaceholder": "Password",
    "signInButton": "Sign in",
    "forgotPassword": "Forgot your password?",
    "register": "Register"
  },
}
```

Listing 3: Start of the `english.json` translation file.

To set translations based on the user's chosen language within the code, the `useTranslations` hook from `next-intl` is used as shown in Listing 4 below.

```
import { useTranslations } from "next-intl";
...
const t = useTranslations("Login");
...
<h2 className="text-center text-3xl font-extrabold text-gray-900">
  {t("title")}
</h2>
```

Listing 4: Example of the translation hook used for the login page title.

⁵⁵<https://next-intl-docs.vercel.app/>

7.2 Backend

As mentioned in subsection 6.3, the backend is the only part of the business logic that is exposed to the internet. This means all interactions from the frontend go through the backend and are authenticated using the User Management Service, which is currently an email and password. If the user is authorized, then the backend delegates the request to the respective microservices to be processed. The backend is a REST API that was built using the Gin framework.⁵⁶ The API can be explored using the generated Swagger⁵⁷ documentation, as depicted in Figure 7.2 below.

The screenshot displays the Swagger UI for the 'SME Connect API' version 1.0. The interface includes a header with the API title and version, a base URL of 'localhost:8081/', and a 'doc.json' link. Below this, there are links for 'API for the SME Connect application', 'SME Connect Support - Website', 'Send email to SME Connect Support', and 'OpenAPI'. An 'Authorize' button is visible on the right. The main content area is titled 'default' and lists several API endpoints with their respective HTTP methods and descriptions:

- POST** /auth/Login Login a user
- GET** /companies/list Get a list of companies
- GET** /companies/search Search for companies
- POST** /company Create a company
- GET** /company/{id} Get a specific company
- PUT** /company/{id} Update a company
- DELETE** /company/{id} Delete a company
- GET** /listing Get all listings

Figure 7.2: Generated Swagger Documentation.

⁵⁶<https://gin-gonic.com/>

⁵⁷<https://swagger.io/>

7.3 Implementation of our Search Approaches

This section discusses how we implemented the three different information retrieval systems (IRS) mentioned in section 4. As the platform has to be able to change between the three information retrieval systems, we also explain how this was accomplished.

7.3.1 Changing between Systems

The key functionality for the platform and the focus of our work is the ability to search and explore the platform and find relevant companies and listings to a user's query. Our implementation has to fulfil the requirement of being able to change between each of the three information retrieval systems (IRS) without having any noticeable changes to the user apart from the searching capability. This requirement is important to evaluate and compare the three different IRS without any external dependencies.

To use multiple information retrieval systems, we must build a consistent but flexible API, which is why we rely on interfaces for consistency. For each service, we define an interface for the functionality that it exposes to the gRPC layer. These interfaces, such as in Listing 5 below, serve as a contract that the different implementations of the three IRS implementations need to fulfil.

```

type ListingDatabaseInterface interface {
    CreateListing(listing *models.DatabaseListing)
        (*models.DatabaseListing, error)
    GetListing(id int64)
        (*models.DatabaseListing, error)
    // shortened ...
}

```

Listing 5: Extract of an interface for listing service

Figure 7.3 below provides a basic overview of how the interface and its implementations are connected.

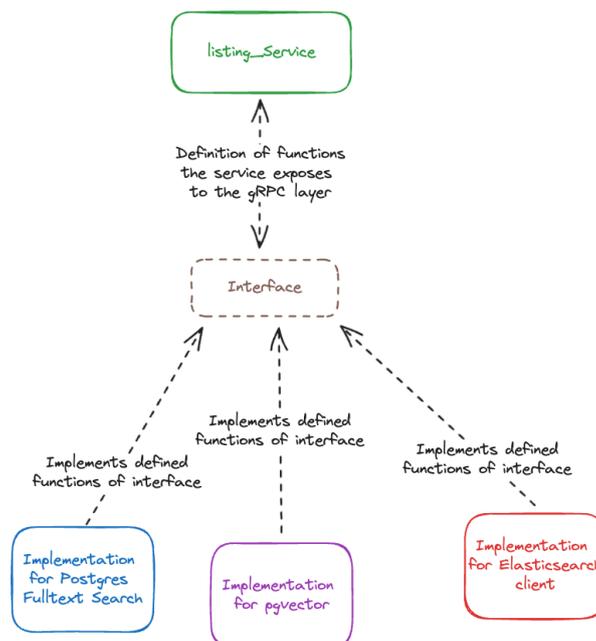


Figure 7.3: Diagram of the interface architecture for the part of the company and listing management service (CLMS).

To differentiate between our three search approaches, we use the environment variable `IRS_MODE`. Internally, the `IRS_MODE` is then mapped and implemented to an enumeration. In the service, we can then change the implementation of the defined interface based on the IRS configuration, *i.e.* `IRS_MODE`. This can also be called the “strategy pattern”.⁵⁸ The strategy pattern, first defined in the book “Design Patterns: Elements of Reusable Object-Oriented Software” [48], is a design pattern in object-oriented programming that fosters flexibility and maintainability by decoupling the selection of specific algorithm from their usage. This approach achieves separation of concerns by separating the “what” (the functionality) from the “how” (the implementation). In Listing 6 below, we illustrate this pattern.

```
func GetListingDatabaseInstance() *ListingDatabaseImplementation {
    IrsMode := utils.GoDotEnvVariable("IRS_MODE")
    if IrsMode == string(types.BASELINE) {
        listingInstance = &ListingDatabaseImplementation{
            databaseConnector: &baseline.
                ListingDatabaseBaseline{},
        }
        return
    } else if ...
        // check for other options
    })
    return listingInstance
}
```

Listing 6: Initialization of the specific IRS implementation.

⁵⁸<https://refactoring.guru/design-patterns/strategy>

7.3.2 PostgreSQL Full Text Search

For the baseline of our work, we selected PostgreSQL Full Text Search for the reasons previously mentioned in section 4.

To implement full-text searching, PostgreSQL provides a parser to create the lexemes from the inserted document; a document being a few words, a sentence, or even a whole paragraph. These lexemes are then a normalized representation of the text. The parsing, in this case lemmatization, includes tokenization, normalization of upper and lower case, removal of stop words and transforming words to their root form, known as stemming.

PostgreSQL then uses this parser with the two data types `tsvector` and `tsquery`. Each document, in the case of our platform, the companies and their listings, are parsed to a `tsvector`, an alphabetically sorted list of lexemes with their position index. These `tsvectors` are then stored within an index that allows for fast retrieval.

When a user performs a search, the search query goes through the same parser but is then represented as a `tsquery`, which is also a list of lexemes. The `tsvector` and `tsquery` types are then used to perform the actual search, using the `@@` operator, which returns a `boolean` for each result. The return value determines whether the two vectors of lexemes match (`true`) or not (`false`). PostgreSQL then automatically orders these results using the `ts_rank` function, which is based on the **relevance** of the search results. An example of the search for a company can be seen in Listing 7 below.

```
var companies []*models.DatabaseCompany
err := db.NewSelect().
    Model(&companies).
    ExcludeColumn("name_tokens", "embeddings", "description_tokens").
    Where(`name_tokens @@ websearch_to_tsquery('english', ?)
        OR description_tokens @@ websearch_to_tsquery('english', ?)`
        , query, query).
    Limit(limit).
    Offset(page * limit).
    Scan(context.Background())
if err != nil {
    return nil, err
}
```

Listing 7: The code for searching for companies. The `name_tokens` and `description_tokens` are the `ts_vectors`.

7.3.3 Semantic Search

Unlike lexical search, where the information retrieval system looks for exact matches of the query words or variants of them as mentioned in subsection 3.2, semantic search can look for matches by understanding the meaning and context of a query and improve the quality of search results with the use of text embeddings.

Generating Embeddings

For semantic search to function, the texts' semantic and contextual similarity first needs to be captured. This is done with an embedding model which creates text embeddings from provided texts. These texts are the query and the document content, in our platform, the title and description of a company or listing. The embedding model then outputs a vector of a specific dimension. Depending on the chosen embedding model, these dimensions can vary dramatically. The output vector is also often referred to as a latent vector because it contains all the latent features of the text. These embeddings are special as they have one key feature. Text embeddings that are close together in the vector space also have a similar contextual meaning. Close together, in the case of vectors, can have multiple meanings as illustrated in the Figure 7.4 below:

- The euclidean or L2 distance between the 2 vectors is very small.
- The inner or dot product between the 2 vectors representing the projection of one vector onto another becomes very large.
- The cosine distance representing the angle between the 2 vectors becomes very small.

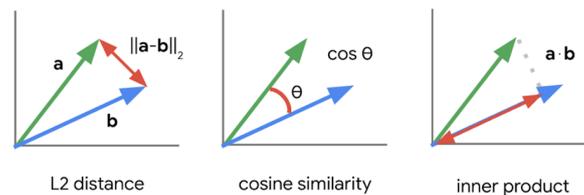


Figure 7.4: Visualization of distance functions between vectors. Image source: [49]

Often, this distance between the vectors is also referred to as “confidence”, as in the task of the retrieval, the distance can be interpreted as how confident the system is that the retrieved vector is correct. If the text embedding is close to the query, then the system is confident that it is the correct one; if it is far away, the system is less confident that the vector is the correct one.

In our platform, the embedding service generates the text embeddings. Initially, we identified the Go package `langchaingo`,⁵⁹ which allowed us to interact with embedding models hosted on Huggingface via their API. Huggingface is the biggest platform for pre-trained models. However, we quickly encountered issues with API rate limits. We then switched to OpenAI’s API but encountered the same issue of rate-limiting when embedding all the companies and listings to populate the platform.

This is when our microservice architecture started to come into its own. As the two previous approaches were unsuccessful, we decided to run an embedding model on our SwitchEngines server. However, we had to ensure the embedding model was small enough to run with our limited resources. Python has the largest community and support for artificial intelligence development, which led us to rebuild the embedding service in Python. Python also has a package for communication via gRPC, so we could still continue using gRPC rather than switching to a REST API approach. We used the sentence-transformer package⁶⁰ to run the embedding model locally on our server. Given a specific sentence-transformer

⁵⁹<https://github.com/tmc/langchaingo>

⁶⁰<https://www.sbert.net/>

model name, the library automatically copies the pre-trained weights/parameters and then builds and starts the model. This allows the embedding model to be easily exchanged in the future if a newer model appears that performs better than our current embedding model or more resources become available. At the time of our work, the best sentence-transformer for retrieval that could run on our server was all-MiniLM-L6-v2. Listing 8 below, shows the setup of the embedding model and the generation of a text embedding.

```
class EmbeddingService(embedding_pb2_grpc.EmbeddingServiceServicer):

    def __init__(self):
        EMBEDDING_MODEL_NAME = os.getenv("EMBEDDING_MODEL_NAME")
        if not EMBEDDING_MODEL_NAME:
            raise EnvironmentError("EMBEDDING_MODEL_NAME
                                   environment variable is not set.")

        # load the model
        self.model = SentenceTransformer(EMBEDDING_MODEL_NAME)

    def CreateSMEEmbedding(self, request, context):
        with torch.no_grad():
            # generate the text embedding
            embedding = self.model.encode(request.content)
        return embedding_pb2.CreateEmbeddingResponse(embedding=embedding)
```

Listing 8: The EmbeddingService without the communication layer.

We can then take all of our companies and listings, create text embeddings for each one, and store them in a database specialized for vectors. When a user sends a search query, this query is also embedded and a nearest neighbour search is performed of the query vector and the stored company and listing vectors. This process is the retrieval stage, as the nearest neighbours of the query are retrieved from the vector database. In our case, we retrieve the 10 nearest neighbours. However, this can be adapted to retrieving any number of nearest neighbours or only retrieving the neighbours within a certain threshold to our query vector. If the euclidean distance is being used to compare vectors, this threshold can be interpreted as retrieving all the neighbours in a certain radius. These retrieved neighbours are the search results for a given query.

pgvector

For the storage of the text embeddings and nearest neighbour search, we use pgvector, an open-source extension for PostgreSQL databases. With the help of this plugin, we can store vectors and compare them using the three different distance functions mentioned in subsection 4.2.

When conducting a search, the vector of the database entry and the vector of the search query need to be compared. The parameters for a search are defined with an interface for each service, as shown in Listing 9 below.

```

type GetListingsResultsBySearchRequest struct {
    Query  string
    Page   int32
    Limit  int32
}

```

Listing 9: The interface defining the search parameters for the listing service in the company and listing management service (CLMS).

In Listing 10 below, the process of performing a search is illustrated. The `searchVector`, is the search query after it has been embedded by the embedding service. The `searchVector` is then used in combination with the cosine distance to perform the search. We use the cosine distance to compare vectors because the embedding models were trained with the cosine distance as part of their loss [50].

```

// for pgvector there are multiple ways to do the search
// 1. using the "<=>" operator, which does cosine distance
// 2. using the "<->" operator, which does euclidean distance
// 3. using the "<#>" operator, which does inner or dot product
err = db.NewSelect().Model(&listings).
    Column("id", "name", "description", "is_need", "company_id").
    Where("is_need = ?", isNeed).
    OrderExpr("embeddings <=> ?", searchVector).
    Limit(limit).
    Offset(page * limit).
    Scan(context.Background())

```

Listing 10: Database query for the vector nearest neighbour search using pgvector.

HNSW Index

By default, pgvector performs exact nearest neighbor searches, this provides more accurate results than approximate nearest neighbour algorithms at the cost of efficiency. To increase the efficiency of our platform, we use the **HNSW** index. The abbreviation HNSW stands for **Hierarchical Navigable Small World**. It is based on the **skip list** data structure ⁶¹ and is used for approximate nearest neighbour searches. The algorithm generates a layered graph. The layers shown in Figure 7.5 allow for faster traversal time, as the graph can now be traversed in three dimensions instead of only two, resulting in faster retrieval of the nearest neighbours [51]. However, because it uses approximate nearest neighbours, there is a slight loss in relevance.

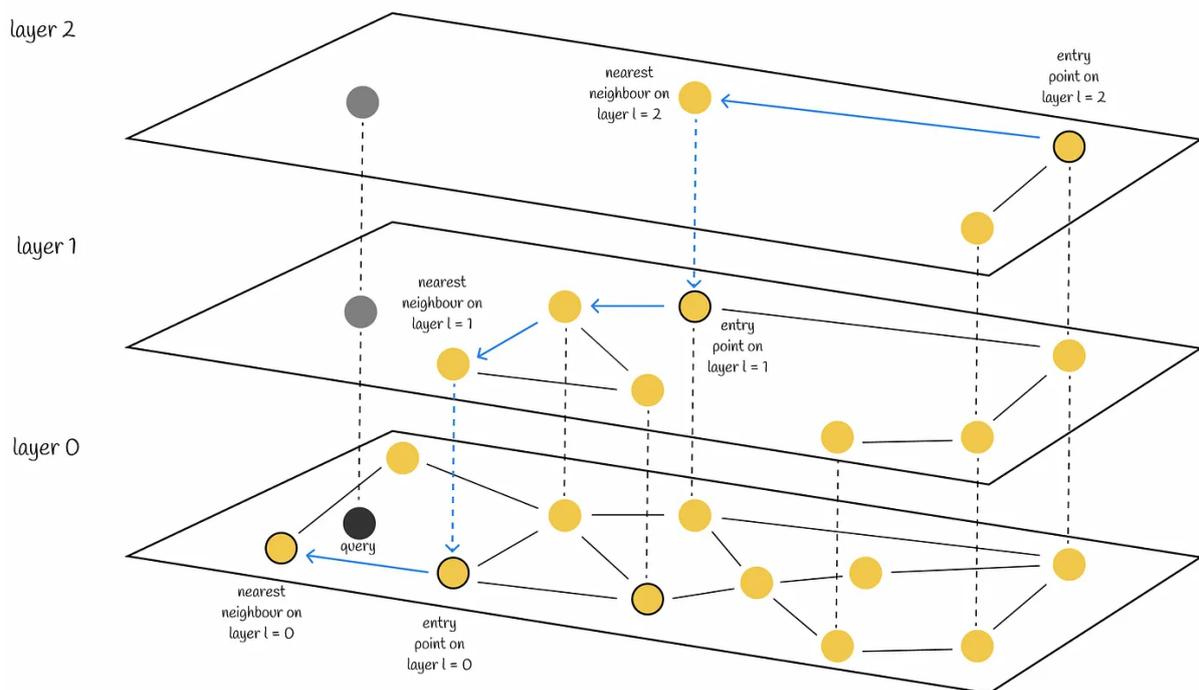


Figure 7.5: Diagram of layered graph traversal. Image source: [51].

⁶¹https://en.wikipedia.org/wiki/Skip_list

To implement the HNSW index in the company and listing management service (CLMS), we define a function that runs after the database table is instantiated, as illustrated in Listing 11 below.

```
func (p*DatabaseListing) AfterCreateTable(ctx context.Context,
query *bun.CreateTableQuery) error {
    _, err := query.DB().NewCreateIndex().
        Model((*DatabaseListing)(nil)).
        Index("listings_embeddings_idx").
        ColumnExpr("embeddings_vector_l2_ops").
        Using("hnsw").
        Exec(ctx)
```

Listing 11: How to configure pgvector to use the HNSW index.

If the searchable range needs to be expanded, fields can be added to the content string that is embedded during insert and update, as already carried out with the names and descriptions as shown in Listing 12 below.

```
embedding, err := client.CreateSMEEmbedding(context.Background(),
    &proto.CreateEmbeddingRequest{
        Content: listing.Name + " " + listing.Description,
    })
if err != nil || embedding.Embedding == nil {
    errMsg :=
        fmt.Errorf("CLMS: Error while calling Embedding Service: %v", err)
    fmt.Println(errMsg)
    return nil, errMsg
}
```

Listing 12: Example of how the name and description of a listing are merged before creating the text embedding.

7.3.4 Elasticsearch

The last search approach that we integrate into our platform is Elasticsearch. The Elasticsearch setup uses the Docker⁶² images provided by the maintainers of the Elasticsearch project. Since a graphical interface is needed to validate the queries and visualize the results, we also deploy Elasticsearch's recommended graphical interface, Kibana⁶³ to the server.

Since Elasticsearch is not the source of truth of our data, we needed an approach to synchronize the data from our PostgreSQL database to the Elasticsearch server efficiently. The solution recommended by Elasticsearch to this problem is Logstash⁶⁴, also a product of the maintainers of Elasticsearch. However, the configuration complexity associated with deploying Logstash as an additional service resulted in us choosing a different approach. Instead, we implemented Debezium,⁶⁵ a dedicated Postgres change data capture (CDC) tool, to efficiently capture and emit database modifications/changelogs. These modifications are then processed by PGSync, another service responsible for indexing them within the Elasticsearch search engine. Figure 7.6 illustrates how the data is synchronized between our PostgreSQL database and Elasticsearch.

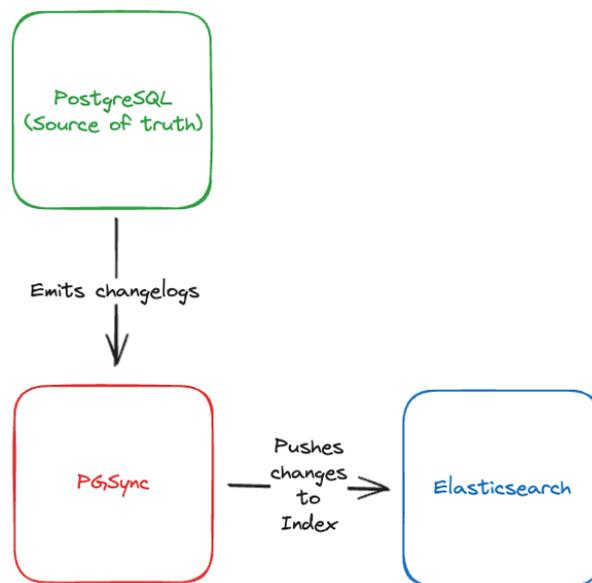


Figure 7.6: Basic Architecture of the sync between the Postgres database and Elasticsearch.

⁶²<https://www.docker.com/>

⁶³<https://www.elastic.co/kibana>

⁶⁴<https://www.elastic.co/logstash>

⁶⁵<https://debezium.io/>

To configure the data synchronization, we need to create a schema utilized by PGSync to map the data between PostgreSQL and Elasticsearch. The schema is defined in the JSON⁶⁶ format utilizing a dedicated file and adheres to the structure in Listing 13 below.

```
[{
  "database": "smeconnect",
  "index": "smeconnect_listings",
  "nodes": {
    "table": "listings",
    "columns":
      ["id", "name", "description", "company_id", "is_need"]
  }
}]
```

Listing 13: PGSync schema for synchronizing PostgreSQL to Elasticsearch.

First, we define the database that we want to sync from, `smeconnect`. Then we define the index that we want to sync to, the index `smeconnect_listings`. Lastly, we define the table that we want to sync from and what columns will be included in the Elasticsearch index, *e.g.* the table `listings` and the columns `id`, `name`, `description`, `company_id` and `is_need`.

When a search request arrives, the company and listing management service (CLMS) acts as a “middle-man”, forwarding the request to Elasticsearch as an HTTP request. Elasticsearch then, just like PostgreSQL FTS, has an index that is used to perform the search and returns all entries that have a match. It then ranks and orders these results using BM25 and returns the results to the CLMS as JSON. The CLMS must fetch the results from the Postgres database if additional information is needed. This allows for Elasticsearch’s index to be as small as possible and keeps the PostgreSQL database as the single source of truth.

⁶⁶<https://www.json.org/json-en.html>

7.4 Code Testing

In this section, we demonstrate our commitment to quality by outlining our code testing process.

7.4.1 Linting

Keeping our code clean and consistent is crucial for development and meeting business standards. Automated tools called linters help us achieve this by checking our code against predefined rules. These rules can be set by the linter provider itself or customized by developers. Linters are specifically chosen for different parts of our codebase. For the TypeScript frontend, we use Next.js's built-in linter. For the Go backend and microservices, we leverage a Go linter. To prioritize developer efficiency, enable frequent committing of code and avoid hindering the local development process, we employ GitHub Actions⁶⁷ that run on our SwitchEngines server to execute these checks. This enables us to ensure that all pull requests into the development or main branch fully conform to our rules.

7.4.2 Frontend

There are two approaches to testing a single-page application such as our frontend. The first approach is that the different user inputs are faked using a headless browser framework like Puppeteer⁶⁸ or Playwright.⁶⁹ These solutions initiate a headless browser window, which means that the user interface is not rendered. Then, tests are performed on the HTML and Javascript code that the website returns. This approach enables developers to simulate keyboard and mouse input and verify if the resulting website output corresponds to the provided user interactions.

The other approach involves unit testing of React components. This method utilizes a testing framework to render the component with predefined data and subsequently compares the resulting output with the anticipated outcome established by the developer. One prominent library facilitating the implementation of such tests is the React Testing Library.⁷⁰

However, the frontend of our platform is still a prototype, primarily serving as a platform for demonstration purposes. Consequently, we opted to prioritize development speed and delay the implementation of comprehensive testing frameworks. Additionally, the reliance on pre-built components from a component library that has been battle-tested by thousands of users further incentivizes this decision, as we anticipate substantial changes to the frontend in the subsequent project phases. This approach prioritizes easy and fast iteration and responsiveness to evolving requirements in the early stages, while we still recognize that a more rigorous setup is needed later.

⁶⁷<https://docs.github.com/en/actions>

⁶⁸<https://pptr.dev/>

⁶⁹<https://playwright.dev/>

⁷⁰<https://testing-library.com/>

7.4.3 Microservices

To ensure the correctness and robustness of the implemented business logic and a solid integration of the different parts, multiple approaches can be chosen: unit testing, divided into black- and white-box testing,⁷¹ and integration testing.

Unit testing requires a large amount of mocking. Mocking utilizes simulated substitutes (mocks) for external dependencies that the code under test interacts with. This isolates the unit, allowing focused testing of its internal logic and facilitating control over the external dependency behaviour (normal operations, errors, etc.). We decided against this approach as the platform does not perform any complex processing or calculations. External dependencies such as PostgreSQL and Elasticsearch take over all of the complex functionality. Since the project mostly involves integrating the different systems, we decided to perform integration testing.

In contrast to unit testing, which examines the functionality of individual code units in isolation, integration testing focuses on verifying how these units work together as a whole. It combines and tests the interaction between several software components, like modules or services. Integration tests specifically target the communication layers to identify potential issues at these crucial interfaces. For instance, they might verify the correct data flow between different modules or guarantee effective communication between various services involved in a system.

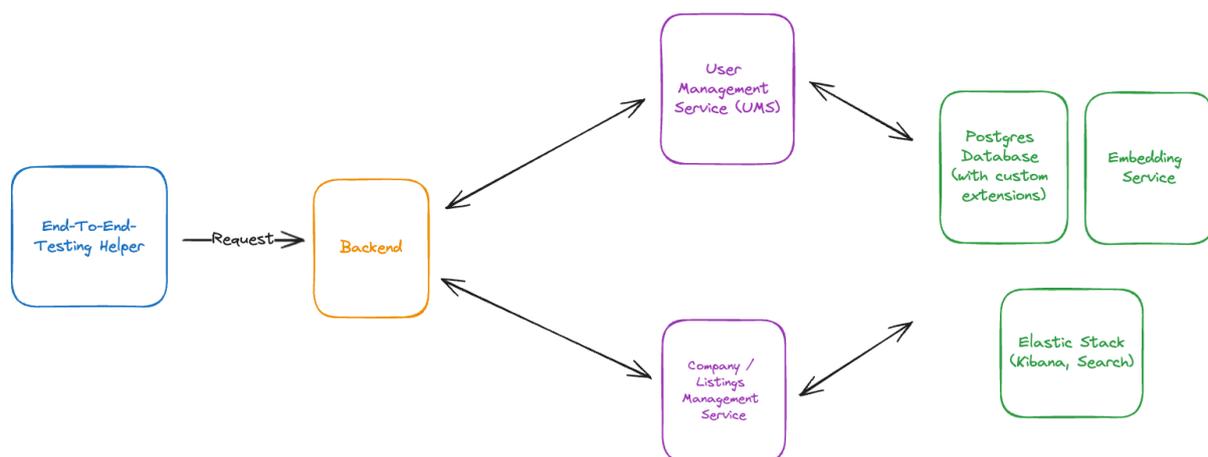


Figure 7.7: Overview of the End-To-End Testing setup.

To perform the tests, we initiate an instance of the whole application, except for the frontend. In this system, we then start an additional helper server that acts as the frontend, as shown in Figure 7.7 above. Using this helper, queries are performed on the backend’s REST API to verify expected results for correct and false input. The helper is written using Typescript and runs with the Bun runtime,⁷² a lightweight, almost drop-in replacement for Node.js.⁷³ Although some features still need to be implemented in Bun, it covers our use case well and reduces the time and computing resources required to run the tests compared to NodeJS. To compare the results, we use Bun’s integrated test framework.⁷⁴

⁷¹Thread on the subject on StackOverflow: <https://stackoverflow.com/a/9893055/12544632>

⁷²<https://bun.sh/>

⁷³<https://nodejs.org/en>

⁷⁴<https://bun.sh/docs/test/writing>

The tests are separated into the domains of the data models they act on, *e.g.* users, listings, companies and sectors. For each domain, we test all the CRUD functions as well as their special constraints, such as ensuring only the specific user can delete his own account. In Listing 14 below, we show such a test.

If a more granular permission system is implemented, the test setup can easily be extended to test those use cases.

```
test("company can be created", async () => {
  // define initial data
  const body = {
    name: "Some Company",
    description:
      "Some Company is a multinational technology company...",
    sector_ids: [3, 4],
    owner_id: 1,
    // shortened ...
  };

  // submit the new data to the backend
  const res = await axios.post(`${backendUrl}/companies`,
    JSON.stringify(body),
    options
  );

  // check whether created input was
  // reflected in the newly created database entry
  expect(res.status).toBe(200);
  expect(res.data.company.name).toBe(body.name);
  expect(res.data.company.description).toBe(body.description);
  // shortened ...
});
```

Listing 14: Example of a test case to test the creation of a company. The test data is anonymised for publication purposes.

The testing process is conducted through Github Actions using the runner that is configured on our SwitchEngines server. We already use these runners to check the code for formatting or conventional errors and ensure that our docker containers start smoothly. As the test results are closely integrated in Github, we can utilize the test results to prevent the merging of changes into our main branch if the tests fail.

Since the docker containers of our platform need time to start up, we defined two methods of ensuring that the testing container will not fail due to our platform not being ready as follows:

1. Each container has a script running that checks if the other containers have started up and are “listening” on their specific ports.
2. We wait an additional time to make sure all the synchronization work can be finished before we perform any operations, as it is crucial to allow the data layer enough time to start up and synchronize the data.

After running the tests, the Github Actions remove any files generated to ensure a fresh setup for the next test run.

8 Evaluation

In this section, we evaluate the three search approaches we implemented in the previous section 7. First, we measure the response time of each search system and compare them. Human evaluators will then assess the relevance of the retrieved results of each search approach in the form of a survey. Finally, an automated evaluation of the search results will be conducted to compare the systems with well-known metrics and a novel approach to finding the overall relevancy of results by using a large language model (LLM) as a judge. It is important to note that the data used for our platform is synthetically generated, which may influence the overall results of our evaluation. The evaluation data and the analysis can be found in the respective Github repository.⁷⁵

8.1 Latency Evaluation

We evaluate and compare the response times of our search approaches to answer research question **RQ2.2** with latency tests. This involves running each microservice within Docker containers specifically designed for the project.⁷⁶ Since production-ready containers are not yet available, we use development containers. To ensure consistent results, all tests ran on a Macbook Pro with an M3 Pro chip and 36 GB of RAM. We kept the device connected to power to avoid performance fluctuations caused by power-saving features. Another reason to perform the tests in a local environment is the networking. If we were to perform the tests with the hosted containers on our SwitchEngine server, we would spam the host's network. This could result in blocking our machine due to DoS protection. Additionally, the results could be impacted by unstable network connections.

We note that the latencies of our tests may vary as we run the development containers. These containers are emulated by the Rosetta layer since the containers were designed for `linux/amd64` [52]. This layer needs to be activated in Docker Desktop, as illustrated in Figure 8.1 below.



Figure 8.1: Screenshot of the Rosetta emulation layer option in Docker Desktop.

⁷⁵<https://github.com/LuciferUchiha/SMEConnectJudge/tree/main>

⁷⁶<https://github.com/raphaelluethy/smeconnect/tree/main/docker>

For the latency tests, we use the `hey`⁷⁷ command line utility within a Bash script,⁷⁸ shown below in Listing 15, as it provides support for parallel requests:

```
#!/bin/bash
SEARCH_QUERIES=(...)
URL='http://localhost:8081/listings/search?query='
FLAGS_2='&isneed=false&page=0&limit=10'
CURRENT_DATE=$(date +%Y-%m-%d)
counter=0
for QUERY in "${SEARCH_QUERIES[@]}"
do
    hey -n 10000 -c 50 -m GET -o csv -H
        $URL$QUERY$FLAGS_2 >> $CURRENT_DATE'_latency_test_results.csv'
done
```

Listing 15: Shortened version of the latency testing script

We select 10 human-written search queries of different lengths and complexity. For all queries, we send a total of 10'000 GET requests from 50 concurrent agents to the `/listings/search` endpoint, resulting in a CSV output as seen in Listing 16 below. We use the `response-time` field for the comparison.

```
response-time,DNS+dialup,DNS,Request-write,
    Response-delay,Response-read,status-code,offset
0.0578,0.0146,0.0010,0.0000,0.0428,0.0004,200,0.0005
0.0639,0.0145,0.0012,0.0000,0.0487,0.0007,200,0.0003
0.0689,0.0143,0.0011,0.0000,0.0542,0.0004,200,0.0005
```

Listing 16: CSV output example

⁷⁷<https://github.com/rakyll/hey>

⁷⁸<https://www.gnu.org/software/bash/>

8.1.1 PostgreSQL FTS

Our latency tests for the baseline information retrieval system with PostgreSQL Full Text Search (FTS), yielded positive results as shown below in Figure 8.2 below. The majority of the response times fall within the very fast 5th percentile, registering at only 0.01 seconds. The overall average response time is also low, at 0.15 seconds. However, the system struggles with frequently occurring generic keywords like “machine learning,” “trading algorithms,” and “cancer.” This is probably because the database has many more matching documents than the other queries and requires more time to determine the most relevant results. Overall, the latency test results align with our expectations, as the search query only needs to be split into lexemes and then the database can perform the search directly.

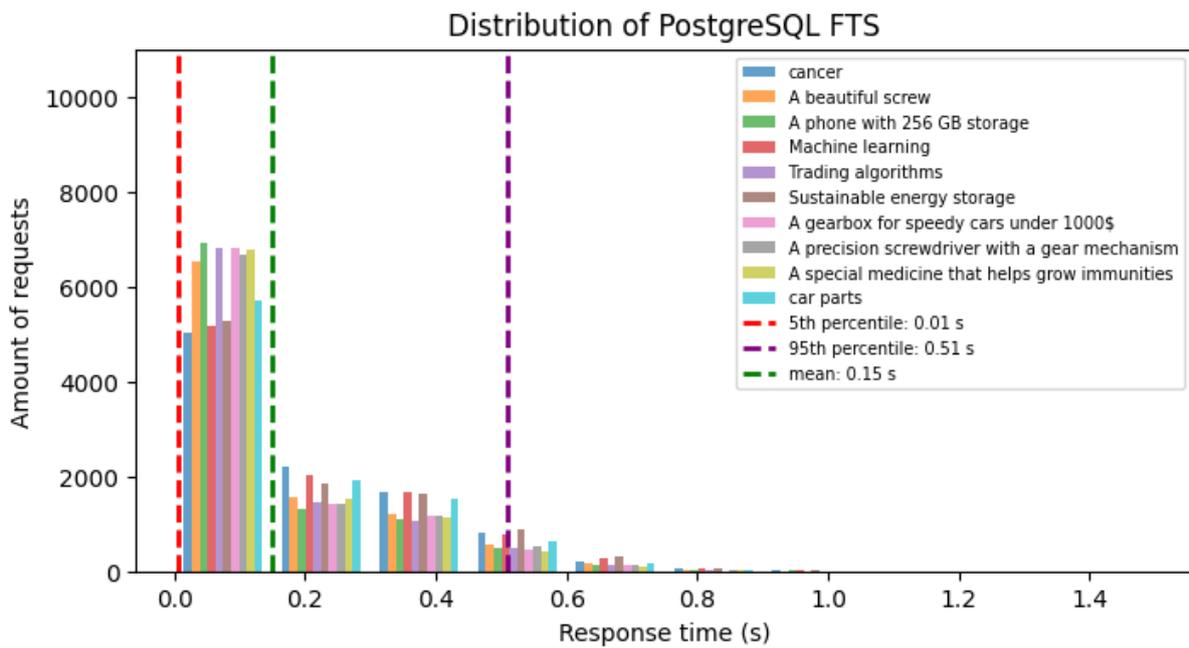


Figure 8.2: Latency test of our baseline, PostgreSQL FTS.

8.1.2 Semantic Search with HNSW Index

Our latency tests of the semantic search with pgvector and the HNSW index also yielded positive results, as shown in Figure 8.3 below. The mean latency of 0.16 seconds significantly exceeded expectations, particularly considering the anticipated overhead from calling the embedding service to generate the text embeddings with a neural network for the search query. This suggests excellent performance from the gRPC layer within our microservice architecture and efficiency of the sentence transformer model. This is a good example of why the additional internal latency of a microservice system can mostly be ignored when using gRPC for communication. The 95th percentile was also lower than expected since we anticipated that more general queries would take longer. Overall, we are impressed with the performance, attributing success partly to the effectiveness of the HNSW index and partly to the gRPC layer.

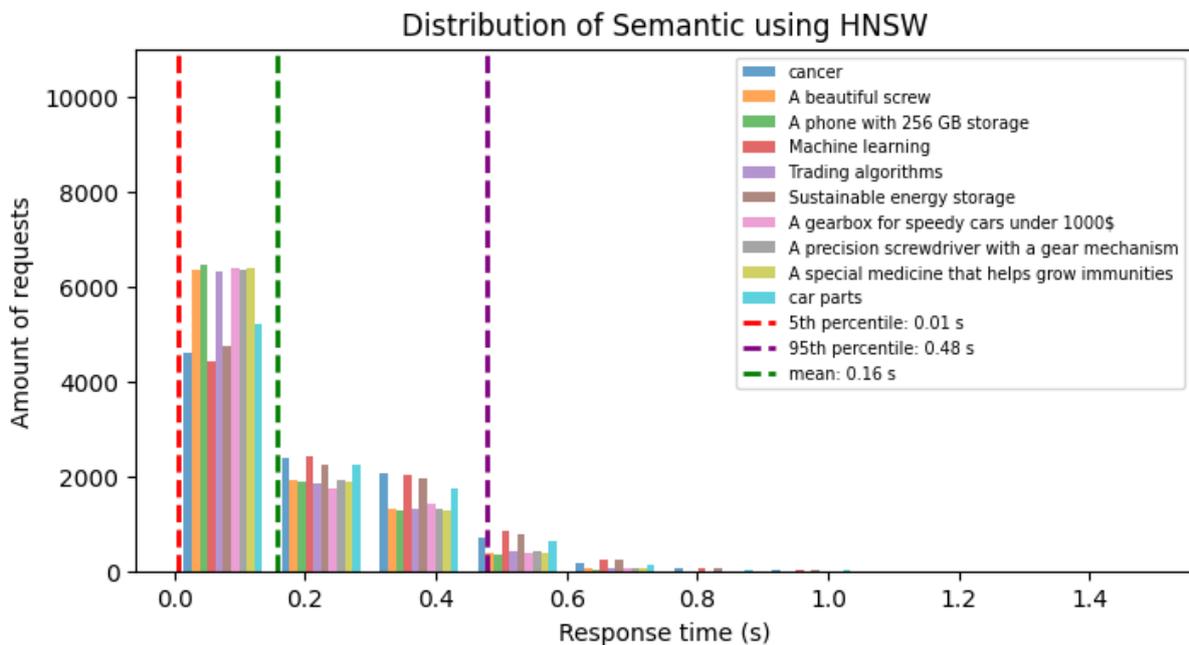


Figure 8.3: Latency test of the semantic search with pgvector and the HSNW index.

8.1.3 Semantic Search without HNSW Index

As shown in Figure 8.4 below, the latency tests of the semantic search with pgvector and without the HNSW index yielded unsatisfactory results. The latencies show a tight cluster between the 5th percentile (0.67 seconds), the mean (0.72 seconds), and the 95th percentile (0.79 seconds) response times, indicating overall system stability. However, the exact nearest neighbour search significantly impacts the latency of the platform. Also, there is a substantial difference in execution time for the query “A phone with 256 GB storage” compared to other queries. The response times of the exact nearest neighbour are worse than expected compared to the HNSW index. We did not expect the HNSW index to have such a significant impact on response times.

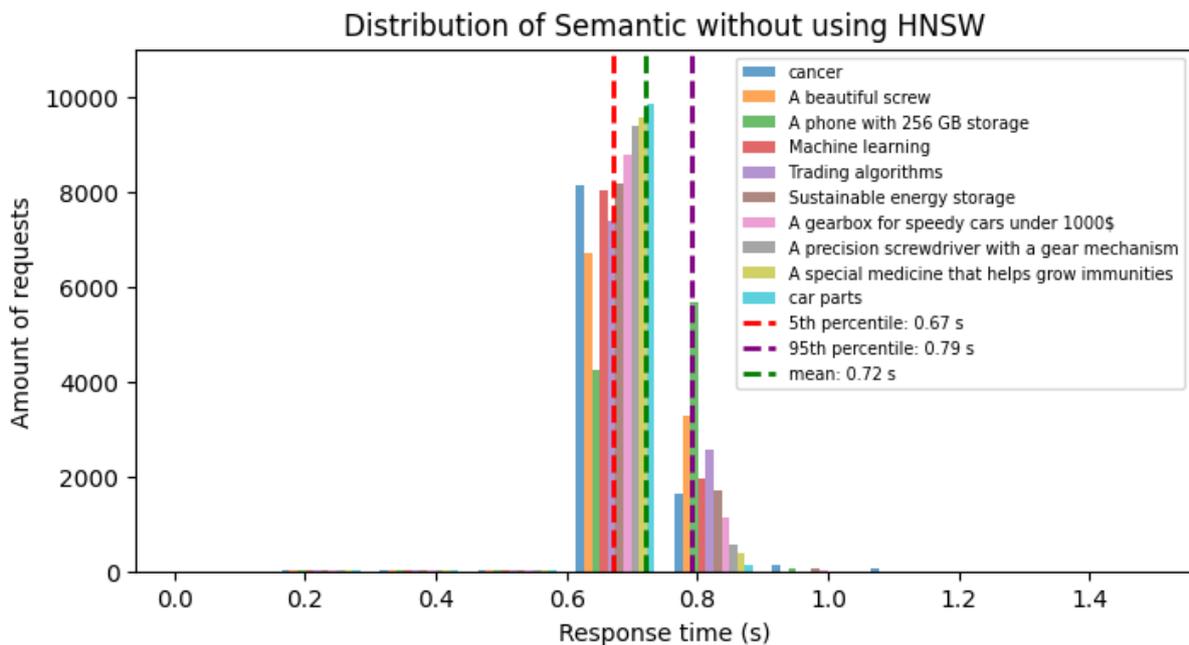


Figure 8.4: Latency test of the semantic search with pgvector and without the HSNW index.

8.1.4 Elasticsearch

Our latency tests revealed strong performance from Elasticsearch. Response times were tightly clustered, with the 5th percentile at 0.02 seconds, the mean at 0.06 seconds, and the 95th percentile at 0.12 seconds, as illustrated in Figure 8.5 below. This aligns with expectations of the industry standard searching solution. It is important to note that only the lexical search capabilities of Elasticsearch were used. The semantic search was not utilized because it is limited to the paid version of Elasticsearch. Consequently, there was no latency overhead from an embedding model. Interestingly, the system exhibited exceptional stability. This is noteworthy as no specific efforts were made to scale or implement parallelism within the Elasticsearch setup. The overall results demonstrate a high-performing and stable search system.

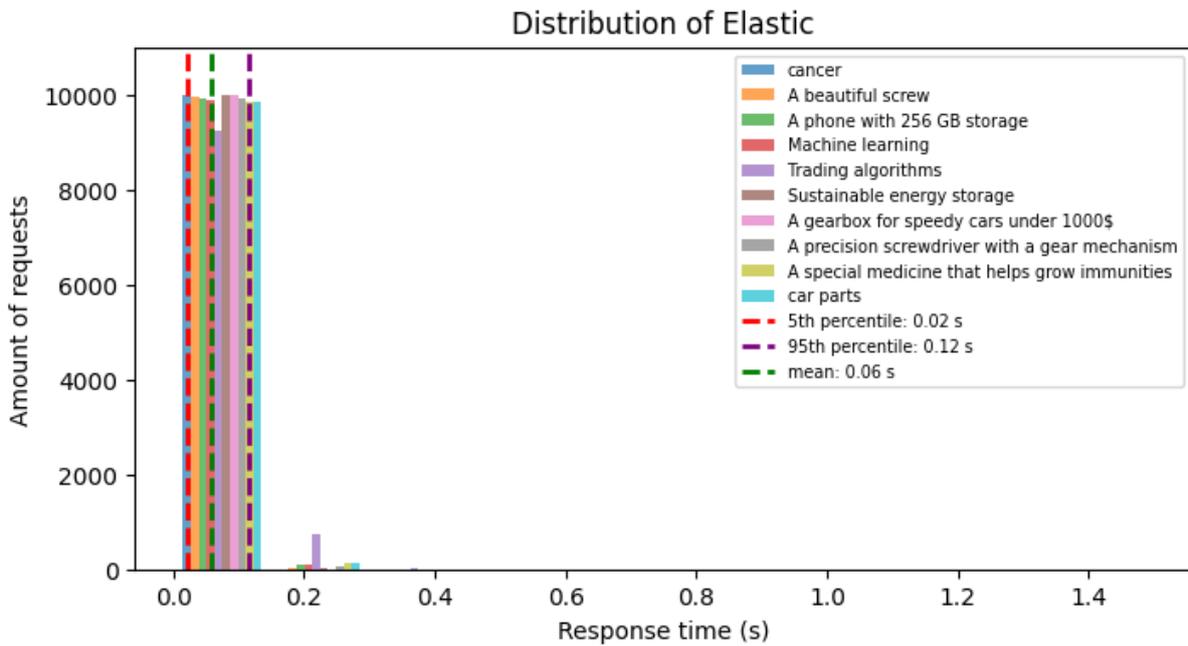


Figure 8.5: Latency Evaluation for Elasticsearch.

8.1.5 Findings

Metric	PostgreSQL FTS	SS with HNSW	SS without HNSW	ES
5th Percentile (s)	0.0065	0.0075	0.6723	0.0217
Mean (s)	0.1501	0.1575	0.7216	0.0594
95th Percentile (s)	0.5089	0.4777	0.7926	0.1157

Table 8.1: Latency comparison of our three search approaches. SS is the short form for Semantic Search. ES is the short form for Elasticsearch.

In conclusion, to answer research question **RQ2.2** of how the response times of the different information retrieval systems compare, we find the following:

- Whilst Postgres Full Text Search (FTS) offers the fastest response times for the vast majority of queries, Elasticsearch is superior regarding both average and worst-case scenarios.
- Semantic search using pgvector requires the HNSW index for acceptable performance.
- Both semantic search with HNSW and Elasticsearch provide interesting search capabilities and maintain low latency, making them good options for our platform.
- Overall, Elasticsearch remains the industry standard due to its stability, speed, and comprehensive feature set. However, semantic search with pgvector using the HNSW index presents a viable alternative depending on the use case.

8.2 Human Evaluation

To evaluate our three approaches to searching and find the answer to research question **RQ2.1** of which approach gives the most relevant results, we conducted A/B tests with 23 test users. We conducted the A/B tests online using a survey. As each information retrieval system (IRS) needs to be compared to one another, we perform three “duals”:

- A vs B
- B vs C
- A vs C

By performing these three duals, we compare each IRS to the others and can define the optimal solution. For the survey, we prepared six queries of varying complexity. We also ensured that we had a query for each of the five business sectors chosen when we generated the data to keep the comparisons fair and diverse. When we defined the six queries, we decided to restrict ourselves to products and services as our qualitative analysis of our synthetic data in subsection 5.3.2 showed that these were of higher quality than needs and desires and we believe that defining search queries for products and services is much clearer for test users than queries for companies. Since we chose to use six search queries and there are three information retrieval systems to compare, we created two queries that each system is more suited to. For example, semantic search prefers more natural flowing language, whereas Postgres’s full-text search prefers keywords. We defined the following six queries:

1. “Cancer” for when a medical research company might look for suppliers of specific equipment used in cancer research.
2. “A phone with 256 GB storage” for when an employer might be looking to bulk purchase smartphones for their employees.
3. “Machine Learning” for when a company developing self-driving cars might be looking for suppliers that specialize in machine learning software.
4. “Trading Algorithms” for when a financial institution might be looking for a company to develop or license a trading algorithm.
5. “Sustainable energy storage” for when a company developing solar panels might be looking for suppliers of batteries or other sustainable energy storage solutions.
6. “A gearbox for speedy cars under 1000\$” for when a go-kart manufacturer might be looking for a supplier of affordable gearboxes.

For each of the three search approaches and each of the six queries, we ran the queries and collected the top 5 search results. We then created the survey with the duals described above and asked the participants to choose which search results they found more relevant. We also allowed participants to choose an option that both search results are equally relevant, as the results of some queries may feel very similar. This resulted in a total of 18 duals to decide which of the implemented information retrieval systems is the best for finding products on a B2B trading platform. An example of an A/B test for the “cancer” query can be seen in Figure 8.6 below. A copy of the full survey can also be found in Appendix C.

* 4. Results 1 of search: "Cancer"

CancerGuard
by Lifeline Biotech

CancerGuard is a groundbreaking cancer detection system developed by Lifeline Biotech. This system uses advanced biotechnologies to detect early signs of various types of cancer, including breast, lung, and colorectal cancers, with unparalleled accuracy. The technology is based on a non-invasive method, making it a comfortable and stress-free solution for patients. In terms of technical...

CancerGuard
by Aegis Biotech

CancerGuard is a revolutionary cancer diagnostic kit that utilizes advanced bio-marker detection technology. The kit is designed to identify the presence of early-stage cancer cells in the body, allowing for timely intervention and treatment. It is a non-invasive test that uses a simple blood sample to detect these bio-markers. The kit has been clinically validated and has a high...

CancerShield
by Sphere Biotech

CancerShield is an innovative oncology drug developed by Sphere Biotech that aims to revolutionize the treatment of various forms of cancer. The drug works by inhibiting the growth of cancer cells and promoting the death of these cells. Unlike traditional chemotherapy, CancerShield is designed to specifically target cancer cells, thereby reducing the side effects associated with treatment. CancerShield is currently...

CancerShield
by Aegis Biotech

CancerShield is an innovative product from Aegis Biotech designed to assist in the early detection of cancer. By utilizing cutting-edge biotechnology, CancerShield can identify the presence of cancerous cells much earlier than traditional methods, thereby increasing the chances of successful treatment. CancerShield is designed for home use, making it a convenient and non-invasive option for...

CancerShield
by Lifeline Biotech

CancerShield is a revolutionary cancer detection system developed by Lifeline Biotech. It uses advanced biotechnologies to identify cancer cells in the body at the earliest possible stage, increasing the chances of successful treatment. The system is based on a complex algorithm that analyzes a wide range of biomarkers to determine the presence of cancer. It is non-invasive and painless, makin...

GenoCure
by BioFlourish

GenoCure is a personalized treatment plan based on a patient's genetic profile. Our team of experts analyze the patient's genome to identify genetic variations that might contribute to the disease. Once identified, a personalized treatment plan is developed to target these variations. GenoCure can be used for a variety of diseases, including cancer, diabetes, and cardiovascular diseases.

BioLuminate Oncology...
by BioLuminate

Our Oncology Therapeutics range is a revolution in cancer treatment. Utilizing breakthroughs in genetic and molecular science, we offer targeted therapy medicines that work by interfering with specific proteins, genes, or processes that contribute to the growth and survival of cancer cells. Our innovative range includes solutions for a variety of cancer types including breast, lung, and colorectal cancers. These...

LuminoTherapy
by BioLuminate

LuminoTherapy is a state-of-the-art cancer treatment that utilizes the power of light to target and eliminate cancer cells. This non-invasive procedure has been proven to be highly effective in curing various types of cancers, with minimal side effects. The therapy involves injecting patients with a photosensitive drug, followed by targeted light radiation. LuminoTherapy has been extensively tested and...

LuminoTherapy
by BioLuminate

LuminoTherapy is BioLuminate's flagship product, a revolutionary new form of therapy for various types of cancer. Utilizing our patented, cutting-edge light technology, LuminoTherapy targets and destroys cancer cells with remarkable precision, minimizing the damage to healthy cells. It's a non-invasive and patient-friendly alternative to traditional chemotherapy and radiation. The therapy can b...

GenomeShield
by Vitality Biosciences

GenomeShield is a cutting-edge genetic editing technology offered by Vitality Biosciences. It is designed to provide highly precise and efficient genetic modifications that can aid in the treatment of a wide array of illnesses. GenomeShield leverages advanced CRISPR-Cas9 technology to perform exact genetic alterations, opening the door to targeted therapies for conditions like cancer, neurodegenerative...

Both search results feel equal

Figure 8.6: An example of an A/B test for the query “cancer”. The semantic search returned the five results at the top. PostgreSQL FTS returned the five results at the bottom.

8.2.1 Participants

We used non-random sampling, specifically convenience sampling, to identify our participants for the survey. This means we specifically targeted people known to us to have experience with B2B trading or experience with information retrieval systems. By selecting the participants this way, we can obtain insights from people familiar with the domain and search functionalities. However, we are aware that there is a potential to create a bias, particularly with a more technically knowledgeable audience, which might not reflect the knowledge of the entire user base.

We initially asked 44 people to participate in our survey. However, we also received feedback from two participants who said they had shared the survey with others they know from the domain. Of these 44 participants, we can see in Figure 8.7 below, that 23 participants completed the survey, resulting in a completion rate of 52%. This high rate is despite the fact that the survey took roughly 20 minutes to complete and was only available in English. Unfortunately, some participants did not take or complete the survey due to the language barrier. However, we could not avoid this issue because our data in the search results is only in English.

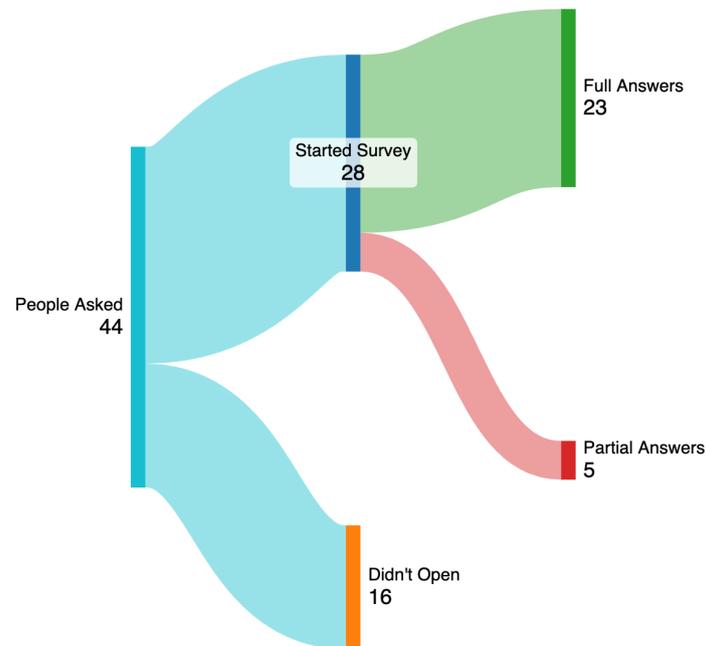


Figure 8.7: Breakdown of the survey participants.

We intentionally did not collect certain demographic data due to privacy concerns, as they can be perceived as intrusive and discourage participation, especially for sensitive demographics such as race, age or ethnicity. We are more interested in the participants' opinions of our search results. However, we did collect data from the participants on the industry in which they are currently working and whether they are currently or have previously been involved in B2B trading, as we believe this could give us some valuable insights.

Of the 23 participants, 11 worked in the IT and software development industry. This was to be expected as it is also the industry in which we know the most people. However, we also had participants from the manufacturing, engineering and energy sectors. A detailed breakdown of the participants' occupational industries in Figure 8.8 is shown below. Of the 23 participants, 56.5% are currently or have previously been involved in B2B trading. This corresponds to our expectations as we targeted people with knowledge in the domains of our work.

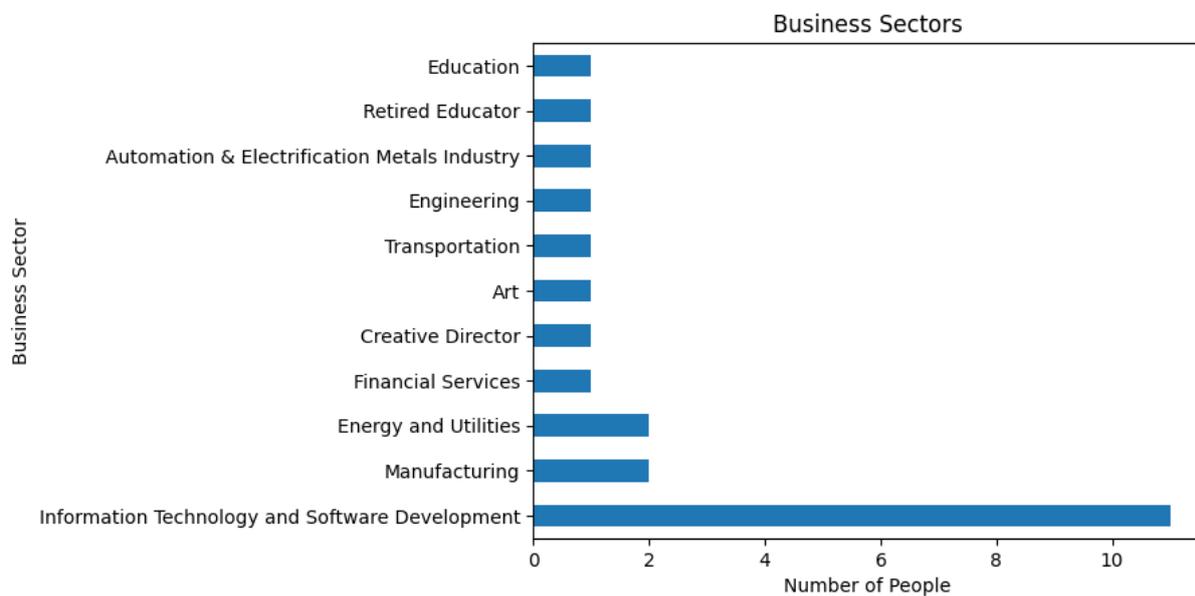


Figure 8.8: The different business sectors of all survey participants.

At the end of the survey, we also asked the participants if they would be willing to participate in a follow-up interview to further aid the future development of the platform. Of the 23 participants, 10 participants would be willing to participate in a follow-up interview, which was higher than we had anticipated. Due to time constraints, we could not conduct any interviews. However, the contact details of potential interviewees have been retained for potential use in the future.

8.2.2 Comparison of Search Results

For each of the six queries, there are three duals, as described previously. Within a dual, a participant has to pick one of three options. They could either choose between two search results, identifying the one they found more relevant or, if they felt that the results were equally relevant, they could choose the third option of “both results feel equally relevant”. We can then count the number of votes/selections each search result receives and, therefore, also the corresponding information retrieval systems.

We define three different voting schemes to ensure the scheme does not impact the results:

- **Only wins:** We only count the votes where a participant specifically chooses a search result.
- **2 Points for wins, 1 point for ties:** If a participant specifically chooses a relevant search result, that information retrieval system receives two votes. If the participant chooses the option of the results being equally relevant, the two points are split equally between the two information retrieval systems; one each.
- **Football system:** This scheme works in the same as the points systems in football leagues; three points for a win, 1 point for a tie.

These votes can be summed up across all six queries to find the participants’ preference of our search approaches and answer research question **RQ2.1** of which IRS gives the most relevant results. Irrespective of the voting scheme used, the semantic search receives the most votes and is therefore the preferred solution. Elasticsearch closely follows the semantic search, and the system with the least number of votes is the PostgreSQL FTS.

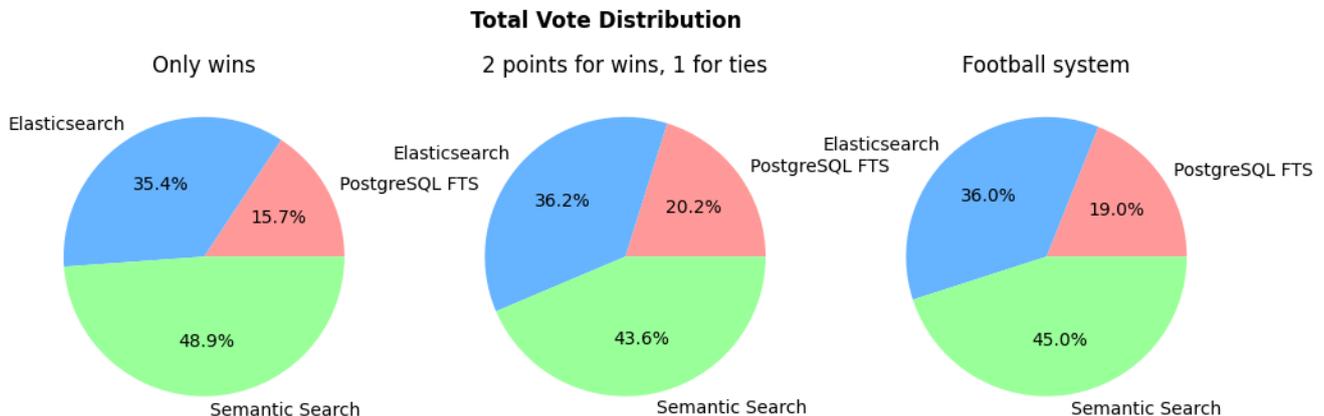


Figure 8.9: The total vote distribution for all three information retrieval systems and voting schemes.

We note that for some of the queries, the PostgreSQL FTS did not return any results. This was also explained to the participants in the survey. However, what is interesting to note is that for these search results where participants had to choose between no search results or potentially inaccurate search results, there was a difference in opinion. About half of the participants would choose the answer with no search results, which was surprising to us. However, we believe this is likely due to human preference as some people would rather have something they can work with, whereas others only want high-quality results.

8.2.3 Comparison with other Search Tools

We aimed to compare the search results generated by our three approaches with search results generated by other typically used search tools to answer research question **RQ1.3**. To achieve this, we included a question at the end of each section of the six queries asking participants if they found the search results “better than search results from other search tools”.

However, setting up the survey this way suffered from two key limitations:

- **Undefined Baseline:** In the survey, we did not specify which “other search tools” participants should consider when evaluating the results provided. This broad scope led to a wide range of interpretations and a potentially unreliable comparison.
- **Lack of Specificity:** We did not define a specific search result participants should compare to. This resulted in variable interpretations and potentially mixed responses.

Despite these limitations, we conducted a more granular analysis of some responses. For example, the query “a phone with 256 GB storage” yielded the following results:

- **PostgreSQL FTS:** No results returned.
- **Elasticsearch:** Irrelevant results.
- **Semantic Search:** Highly relevant results.

This spread of results can also be seen in the participants’ feedback illustrated in Figure 8.10 below. Negative responses most likely stemmed from the PostgreSQL FTS results. In contrast, average and positive responses might refer to the results from Elasticsearch or semantic search, depending on the participants’ familiarity with semantic search. However, we can see in our analysis of other B2B trading platforms in subsection 3.1 that participants are not necessarily familiar with search tools that have contextual understanding.

To improve future evaluations of the platform, we propose that a clear set of competitor search tools and comparable results should be considered. We also suggest that participants be provided with sample results from our platform and the competing tools to compare between. This allows for a more focused and accurate comparison.

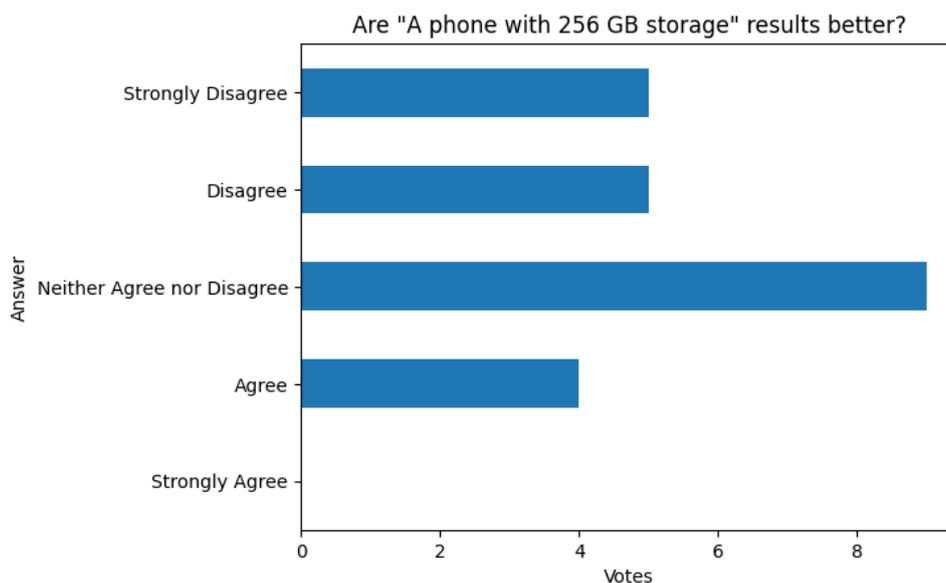


Figure 8.10: Analysis of whether participants found our search results for the query “a phone with 256 GB storage” better than other search tools.

8.2.4 Findings

After performing the survey of our three search approaches with 23 participants, we summarize our findings and answers for our research questions **RQ2.1** of which information retrieval system gives the most relevant results and **RQ1.3** on how our search results compare to other search tools below:

- We found that semantic search is the participants' preferred information retrieval system, receiving 43.6% to 48.9% of the votes depending on the voting scheme used. This preference comes from the capability of finding accurate search results for natural language queries such as "a phone with 256 GB storage".
- Elasticsearch is a very close competitor to the semantic search, receiving 35.4% to 36.2% of the votes depending on the voting scheme used. This search approach especially gains votes when precise keywords are used whilst still being able to also gain votes for handling more complex search queries.
- Unfortunately, PostgreSQL FTS clearly does not meet the search expectation requirements that users have of a search platform.
- Interestingly, on our evaluation of the survey results, there was a split of opinion on whether no results should be returned rather than inaccurate ones. This should be considered when further developing the platform. A possible solution to this issue could be to display search recommendations, general product categories or alternative queries if no results could be found.
- Despite our survey not being set up ideally to compare our search results with those of other search tools, we find that our results are comparable, if not better, for complex natural language queries. We also find that, to have a fair comparison between different B2B trading platforms, a common dataset between all platforms would be required, which poses a great challenge.

By analyzing for each of the three search approaches to which "opponent" they lost the most votes in Table 8.2 below, we can show how close the participants' preference of Elasticsearch and semantic search is. With Elasticsearch losing the most votes to semantic search and both systems winning the most votes from duals against PostgreSQL FTS.

Lost votes against	PostgreSQL FTS	Semantic Search	Elasticsearch
PostgreSQL FTS	-	22	24
Semantic Search	89	-	60
Elasticsearch	55	24	-

Table 8.2: The amount of votes that the bold search approach above lost to the search approach on the left.

8.3 Automated Evaluation

To mitigate the possibility of potential bias in our survey, we incorporate automated evaluation as part of our comparison of the approaches. This evaluation leverages established information retrieval metrics while also using large language models (LLMs) in a novel way to assess and judge how relevant a search result is to a given query. This not only provides valuable insights into our three search approaches but also lays the groundwork for effortless evaluation and comparison of potential future solutions to the current search approaches.

8.3.1 Test Dataset

To automatically evaluate each of our search approaches' performance, we constructed a test dataset of 80 queries that combines the strengths of both human-written and AI-generated queries. This approach offers several advantages.

First, we selected a representative random subsample of 40 companies and 40 products from our original dataset. George Rowlands, Raphael Lüthy and Nitish Patkar then independently wrote queries based on a company or product description for half of these samples *i.e.* a total of 40 queries. By having three different people write queries, we can ensure a variety of query styles, phrasings, and complexity, reflecting how real users might interact with the platform.

Whilst 40 human-written queries provide a solid foundation, we further expanded the test set by leveraging the power of OpenAI's `gpt-4-Turbo` LLM to generate an additional 40 queries. This brings a significant amount of data into the test set efficiently, allowing for a broader range of potential user interactions and analysing the information retrieval systems' ability to handle diverse query formats.

Importantly, for each query in the test set, we also have access to the expected result, allowing us to measure the accuracy of each information retrieval system in retrieving the desired information.

8.3.2 Metrics

Having established a comprehensive test dataset, we now look at the specific metrics employed to evaluate and compare our three search approaches:

Zero result rate (ZRR): This metric indicates the number of queries for which the IRS fails to return any results. We set each IRS to return all results that are relevant no matter by how much. Therefore, this metric also shows the rate at which a company or product will be missed entirely for the given query. Ideally, this metric should be zero, signifying the systems' ability to identify at least some potential matches for every query.

Misses: Since our system retrieves all potential results, this metric is inherently equivalent to the ZRR multiplied by the total number of queries *i.e.* 80 (40 human-written and 40 synthetically generated). However, it is important to acknowledge this metric for transparency.

Rank: This metric shows the position of the expected result within the ranked list of the search results. A lower rank signifies a more relevant result. Ideally, all results would appear at rank 1, the first position. To be more precise, we will focus on the mean and median rank of each search solution. The mean rank tells us the average position of the expected result across all queries. The median rank is a metric less influenced by outliers compared to the mean, which can be skewed by a few very high or low ranks.

Hit rate at value K: The hit rate is the proportion of queries for which the expected result appears if only value K results would be returned *i.e.* the rate at which the expected result is within the top K positions of the ranked results. We calculate hit rates for various values of K *e.g.* 1, 3, 5, 10, 20, 50, 100. The higher the rate, the better. Intrinsicly, if K grows, then so does the hit rate.

Mean Reciprocal Rank (MRR): This metric calculates the average reciprocal of the rank for each query's expected result, the reciprocal being the multiplicative inverse *i.e.* the reciprocal of x is $\frac{1}{x}$. Compared to the mean rank, the MRR is less sensitive to negative outliers as it prioritizes the position of the expected result in the ranking *i.e.* it gives more weight to results appearing at the top compared to those appearing lower. An MRR closer to 1 signifies a higher overall performance, indicating that the expected results tend to appear at higher ranks on average. We also calculate the MRR@5, which means we only calculate the MRR for the queries where the expected result was in the top five results. This gives a more focused view of the search results, as the first five results are the most important.

We note that for the semantic search, we compared the search results using the exact nearest neighbour search (NN) with the approximate nearest neighbour search (ANN) using the HNSW index. For our dataset, both configurations return the exact same search results. However, this may not be true for larger and more diverse datasets. Therefore, when comparing the semantic search with our other search approaches, we use the HNSW index as it significantly improves the response times at no cost to search result accuracy.

IR System	ZRR	Misses	Mean Rank	Median Rank	MRR	MRR@5
PostgreSQL FTS	43.75%	35	2.28	1	0.39	0.41
Elasticsearch	2.5%	2	4.06	2	0.59	0.72
Semantic Search	0%	0	10.39	5	0.37	0.61

Table 8.3: Results of the automated evaluation using all 80 queries for each of the three search approaches.

Setting out the results of each metric in the Table 8.3 above enables us to draw the following conclusions:

- The PostgreSQL FTS misses a lot of results due to not returning any results if no matching keywords are found. However, when it does find a result, then this result is highly accurate. We believe this is due to its strong reliance on keyword searching.
- The semantic search never misses any results. However, this also leads to a lot of expected search results having a very low rank, clearly bringing down its average performance. This can be seen in the difference between the mean and the median rank. If we only focus on the results that have a hit in the top 5 with the MRR@5, we can see the score drastically increases. This can also be seen in Figure 8.11 below, with the semantic search result ranks having a wider distribution than the other information retrieval systems.
- Elasticsearch provides a combination of both of the positives from the other two approaches. It is very accurate and does not have a lot of misses, making it a very good search solution for the platform.

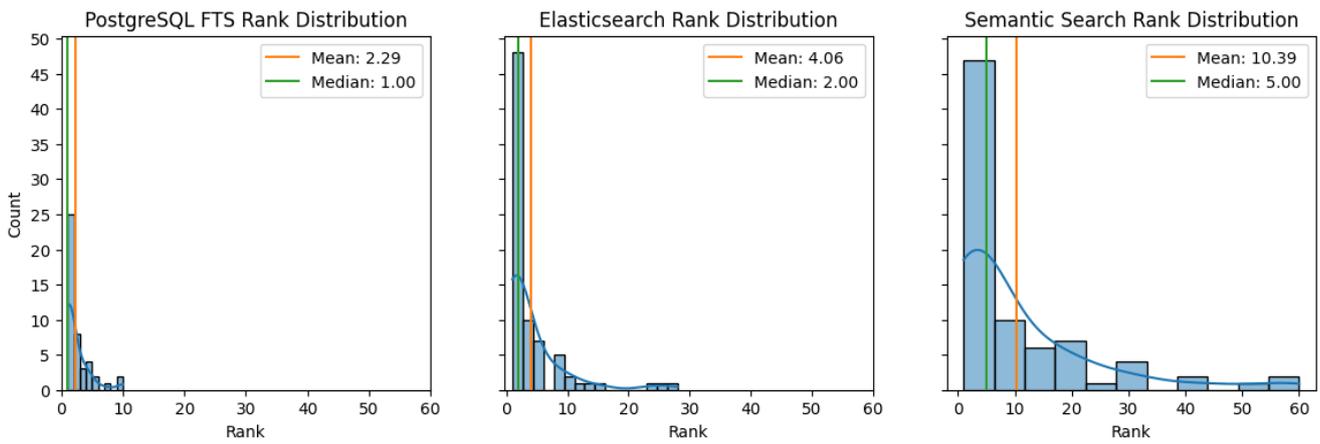


Figure 8.11: The distribution of where the expected result is found.

In Figure 8.12 below, we show how the hit rate increases with the number of returned results. We can clearly see that PostgreSQL FTS quickly reaches its limits at 10 results. However, this could be due to the size of our dataset. The high zero result rate (ZRR) holds PostgreSQL FTS back from breaking the 60% mark. We also see that the Elasticsearch hit rate has very consistent growth in relation to the number of search results retrieved. The semantic search surpasses the PostgreSQL FTS if more than three search results are considered. It also catches up to the elastic search with more results, ultimately overtaking it due to Elasticsearch having some misses.

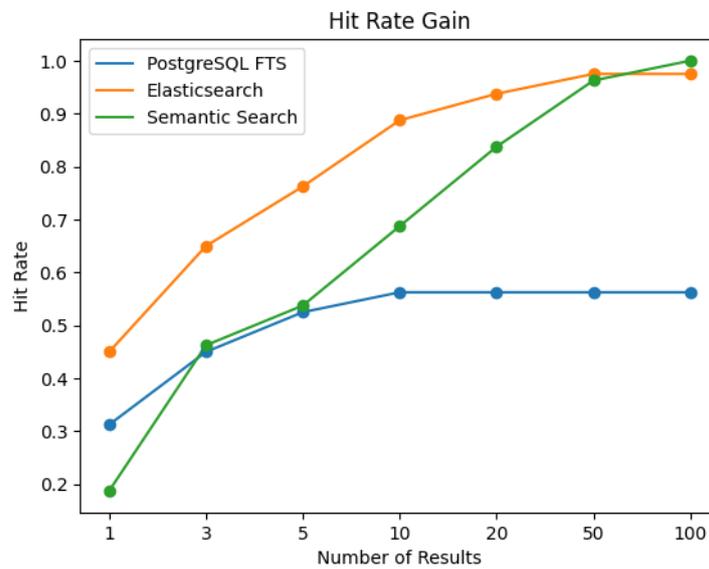


Figure 8.12: The hit rate gain when more search results are considered.

8.4 Human Queries vs LLM Queries

We also investigate the difference between the human-written queries and the queries written by a large language model (LLM) in our test dataset. Table 8.4 below shows that the LLM queries outperform or are equal to the human-written queries for all approaches and metrics. This suggests that LLMs excel at extracting key information from data and translating it into precise search queries. However, on average, the LLM queries were ten characters (approximately two words) longer. While this increased length might contribute to their accuracy, it also raises the question of whether these queries could be optimized for conciseness while maintaining effectiveness.

IR System	ZRR	Misses	Mean Rank	Median Rank	MRR	MRR@5
PostgreSQL FTS Human	55%	22	2.61	1	0.30	0.31
PostgreSQL FTS LLM	32.5%	13	2.07	1	0.49	0.51
Elasticsearch Human	5%	2	4.82	2	0.50	0.64
Elasticsearch LLM	0%	0	3.35	1	0.68	0.80
Semantic Search Human	0%	0	12.48	5	0.35	0.60
Semantic Search LLM	0%	0	8.3	4	0.39	0.62

Table 8.4: Results of the automated evaluation split between the human-written queries and the queries the LLM generated.

8.4.1 LLMs as Judges

In addition to the more traditional metric-based evaluation, we also introduce using large language models (LLMs) as “judges” to assess the relevance of search results. LLMs have more recently gained a lot of popularity as judges. LLMs are uniquely suited for this task due to their ability to understand the intricacies of natural language, context, and search intent. Our novel approach to using LLMs draws inspiration from the concept of context relevance in Retrieval-Augmented Generation (RAG) systems. Similar to how contextually relevant chunks in RAG systems are measured for their relevancy with the user’s query and conversation, we propose analyzing our three search approaches’ top five search results based on their relevance to the specific search query. To mitigate any potential biases of the LLM, we propose that each of the top five search results be judged by the LLM three separate times. The average of these three assessments then determines the final “LLM relevance score” for each result. The prompt we used for “LLM relevance score” can be found in Appendix A. The distribution of these LLM relevancy scores can be seen in Figure 8.13 below.

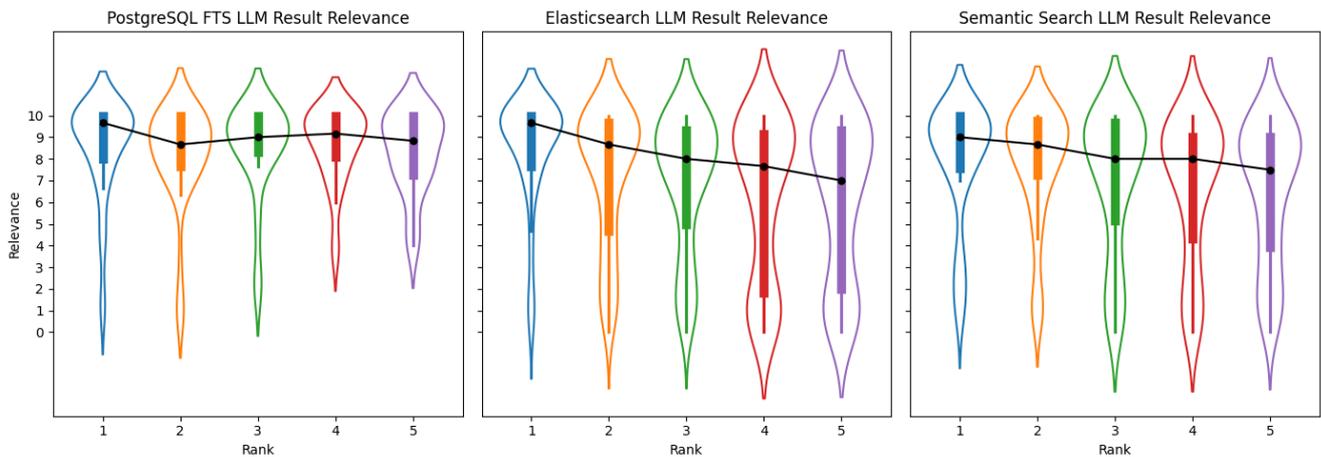


Figure 8.13: Violin plots showing the distribution and median of the LLM relevance score for the top five search results of each of our search approaches, *i.e.* the top-ranked search results. In the centre of each violin is a boxplot, with the black dot being the median value of the data and the box forming the interquartile range of the data. The data range and distribution can also be interpreted from the violin’s shape (a histogram on its side). A wide area showing where data points are more frequent.

Across all three approaches, relevancy decreases with each lower-ranked result, which is to be expected. This shows that our three search approaches return relevant results and that our LLM relevancy score is a valid metric for comparing information retrieval systems.

It is clearly visible that PostgreSQL FTS is very accurate and that the LLM judge is very sure about the relevancy, especially for the first four results. However, we note that PostgreSQL FTS often returns zero results, which are omitted from this graph. The value distributions for the semantic search and Elasticsearch are as expected. However, it is interesting to note that Elasticsearch’s 4th and 5th search results significantly increase the score distribution.

We also analyse the standard deviation of the LLM's relevance score to give us insights into the consistency of the scores. With an average standard deviation of 0.35, meaning the scores for the same query and result tend to differ on average by 0.35 points, we feel confident that the LLM's assessments are consistent and reliable. In Figure 8.14 below, we show the distribution of the standard deviation values for the LLM relevancy scores.

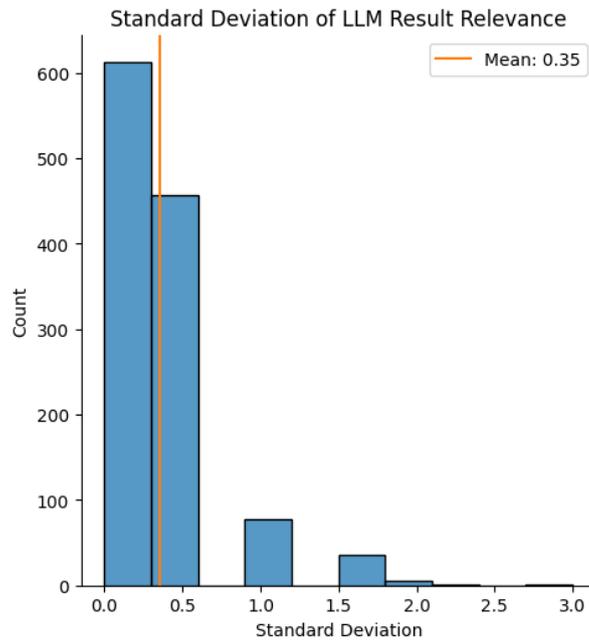


Figure 8.14: A histogram of the standard deviation values of all LLM relevancy scores.

8.4.2 Findings

After conducting multiple automated evaluations and comparisons to answer our research question **RQ2.1** of which information retrieval system gives the most relevant results in a neutral and reusable way we summarize our findings of our automated evaluation below:

- We found that search queries generated by Large Language Models (LLMs) demonstrably outperform their human-written counterparts in retrieving relevant information. This is a significant finding, as it shows the potential of using LLMs to transform and augment user search queries to improve information retrieval. However, we note that the synthetically generated queries were, on average, 10 characters longer than the human-written queries.
- Our evaluation yielded promising results regarding the potential of using Large Language Models (LLMs) as judges of text relevancy. Notably, LLM exhibited high consistency and agreement in its evaluations. The average score deviation for the same question was only 0.35 on a scale from 0 to 10. Therefore, we believe that our LLM relevancy score of a result compared to its query is interpretable and could prove useful in future work in the field of information retrieval.
- Our evaluation exposes a key weakness of PostgreSQL FTS. With a zero result rate (ZRR) of 43.75%, it misses a large amount of results, making it an unsuitable search approach for our platform. This weakness is due to PostgreSQL FTS not being able to handle synonymous words in the query that do not occur in the descriptions. However, we note that if it does return a result, the result is very accurate, with the expected result being on average within the first 3 results.
- Elasticsearch clearly surpasses PostgreSQL FTS by addressing the previously observed weaknesses of misses, while simultaneously maintaining a high search accuracy. This is evident by its superior performance metrics, achieving the highest Mean Reciprocal Rank (MRR) of 0.59 and the highest MRR@5 of 0.72. This, coupled with the low latency observed in our previous evaluation in subsection 8.1, along with Elasticsearch's established reputation for scalability, makes it a very promising solution for our platform.
- In our evaluation, we found that semantic search showed promise, exceeding PostgreSQL FTS in MRR@5 but falling short of Elasticsearch's accuracy. This is likely due to the efficient yet limited embedding model and synthetic training data. Future work using larger embedding models like Hugging Face's Leaderboard's ⁷⁹ top performers could improve our semantic search substantially. In the following section on future work, we define some additional improvements for semantic search to improve accuracy, such as using a cross-encoder to rerank the results [53] or transform queries using large language models [54] to take advantage of their contextual understanding and test if the transformed human-written queries can match the performance found with synthetically generated queries.

⁷⁹<https://huggingface.co/spaces/mteb/leaderboard>

8.5 Impact of Data Quality

When choosing an information retrieval system for a platform, the nature and quality of the data and desired search behaviour must be considered. Postgres Full Text Search (FTS) prioritizes keyword matching, favouring shorter texts densely packed with relevant, qualitative terms. In comparison, semantic search excels with longer natural language texts, leveraging text embeddings to account for synonyms, typographical errors, and contextual meaning. Elasticsearch falls between these two, offering some understanding of typographical errors and word variations while emphasizing the importance and quality of keywords within the document. The optimal choice of an information retrieval system for a platform depends on whether to prioritize exact keyword matches in potentially shorter content, semantic understanding in longer texts, or a balance between the two with some tolerance for typographical errors.

8.5.1 Findings

We can conclude our findings for research question **RQ2.3** that the data quality is important when choosing the search's underlying technology and, therefore, also impacts the resulting search results. This, and the impact of the synthetically generated data, can be visualized for the semantic search after projecting the text embeddings to two dimensions as in Figure 8.15 below. Each of the clusters is due to the business sectors. However, we can see that the clusters are very tight. Because the clusters are so tight, the semantic search can have difficulty choosing the correct result. These tight clusters can be interpreted as the synthetically generated descriptions by the LLM being more general and using commonly used words rather than precisely pinpointable keywords. In comparison, human-written descriptions are of higher quality and more diverse, especially if the people writing the descriptions use techniques to optimize their descriptions to appear more frequently in searches, so-called search engine optimization (SEO).

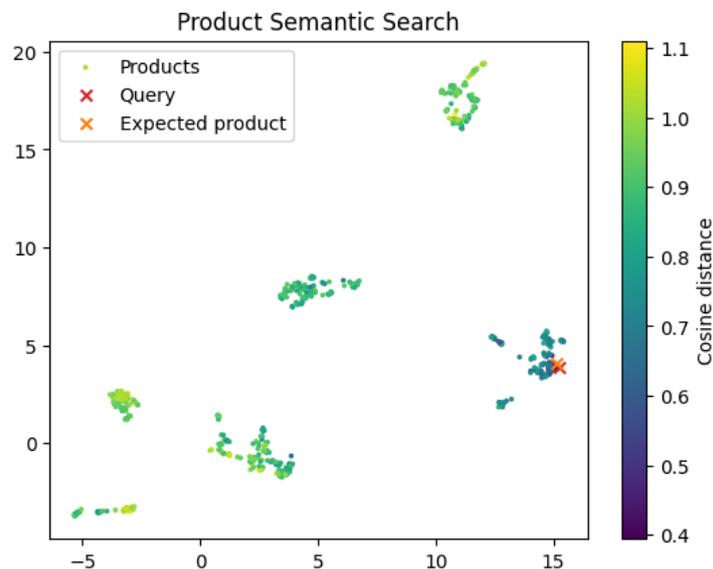


Figure 8.15: The cosine distance of search results to the search query is very tightly packed together, making the choice difficult.

9 Conclusion and Future Work

9.1 Conclusion

In this thesis, we compared three search approaches for a B2B trading platform: PostgreSQL Full Text Search, semantic search with pgvector and Elasticsearch. Our work mainly focused on evaluating these three information retrieval systems for our platform and its specific use case. In addition to the comparison, we planned and built the foundation for the platform, from architecture design to implementing the various microservices and information retrieval systems.

From our extensive evaluation and comparisons, we conclude our research questions as set out below:

RQ1.1: What features can aid a user during exploration?

Our analysis of five existing B2B trading platforms in subsection 3.1 and the general state-of-the-art in searching revealed a significant gap in the B2B trading platform's search functionality compared to the current state-of-the-art. This highlights the potential for improved user experience and search accuracy. Incorporating features such as multilingual support, conversational searching, search history, search recommendations, and query auto-completion can improve the user experience during exploration.

RQ1.2: How do B2B trading platforms' search features compare?

Our comparison of the five B2B trading platforms revealed that there is a gap to be filled in search functionalities, also when comparing them to the general state-of-the-art in searching. No platforms support contextual and semantic search. While some platforms suggest searches, automatic completion and offer multilingual support, no platform completely covers all features. By integrating keyword highlighting, filter suggestions based on intent and covering all of the features mentioned above, a B2B trading platform that surpasses those in the current market could be created.

RQ1.3: How do users perceive the search results of our platform compared to those of similar platforms?

We demonstrated in subsection 8.2.3, that even though our survey was not set up optimally to compare our search results with those of other search tools, with in-depth analysis, we can come to the conclusion that our results are comparable to, and in some cases even better than other B2B trading platforms. Our search approaches are especially impressive compared to other search tools when given natural language queries. However, we concluded that to conduct a fair comparison between different B2B trading platforms, a common dataset or equal examples should be used.

RQ2.1: Which of our implemented IRS gives the most relevant results when exploring companies and their products?

Based on the findings from our survey in subsection 8.2.4 and the automated evaluation in subsection 8.4.2, we conclude that Elasticsearch and semantic search are the most promising information retrieval systems (IRS) for our platform. Elasticsearch excels in search accuracy and scalability whilst still being able to handle complex queries. This makes it the ideal choice for larger datasets and a wide range of queries. Semantic search, while demonstrating strong search accuracy after excluding outliers, may improve even further with non-synthetic data and minor refinements such as a larger embedding model and query transformations. It excels at long and complex natural language queries where understanding context is important. In contrast, PostgreSQL FTS, while retrieving highly relevant results, exhibits a critical flaw in failing to return results for nearly half of the queries in the test dataset. However, is still a viable option if the searches only use simple keywords in short texts.

Therefore, we believe a hybrid approach that leverages both Elasticsearch and semantic search, or utilizing the paid version of Elasticsearch, is the most optimal solution for our B2B trading platform.

RQ2.2: How do the response times of the different IRS compare?

In subsection 8.1.5, we performed latency tests for our three search approaches. Our tests revealed that Elasticsearch delivered the most consistent and reliable performance across all search queries. While PostgreSQL FTS also displayed favourable response times, its performance decreased noticeably for queries involving frequent keywords. This suggests potential limitations in PostgreSQL FTS's scalability. Furthermore, our findings strongly support the adoption of the HNSW index to enable efficient semantic search within our platform. Searches executed without the HNSW index take, on average, almost five times longer.

RQ2.3: How do the different IRS behave when subjected to various qualities of company descriptions?

We discuss in subsection 8.5.1 how the quality of text can impact the search accuracy of information retrieval systems (IRS). We come to the conclusion that our approaches underperform when commonly occurring words or very similar descriptions are being compared. If the IRS is a lexical search, it can also struggle with synonyms as it can not understand the meanings of a word, unlike a semantic search.

We also reflected on the development process of the B2B trading platform and the integration of our three search approaches with the following observations:

- While implementing PostgreSQL FTS and the semantic search with pgvector was relatively straightforward, integrating these systems into the platform required significant effort.
- We were surprised that our initial approaches to using HuggingFace's API and OpenAI's API for the embedding service did not work due to rate-limiting. However, in hindsight, we believe that having a self-hosted embedding model, whilst requiring a more powerful server, is a much better option as it can be significantly cheaper and allows for more control and flexibility over the model itself.
- Our decision to use Go for microservices paired with gRPC proved highly effective. Go's performance and approachable syntax made it an excellent choice. Overall, we are very satisfied with the chosen technology stack for the platform.

In conclusion, this thesis provides valuable insights into information retrieval systems for B2B trading platforms by evaluating PostgreSQL Full Text Search, semantic search with pgvector, and Elasticsearch. Our work builds the foundation for a new B2B trading platform while also providing suggestions on our platform's search functionalities.

9.2 Learnings

Our thesis was a valuable learning experience for us both. We delved into information retrieval, a new area for us. Additionally, we built a platform using microservices with Go, a programming language we specifically learned for this project. Finally, we mastered the intricacies of writing and structuring a thesis with LaTeX, a skill that will prove beneficial to our academic future.

9.3 Future Work

In the following section, we discuss future research or development that could further improve the platform or our three search approaches.

9.3.1 Matchmaking

A potentially interesting feature for our B2B trading platform could be to explicitly and automatically match companies with products or companies with other companies. These matches could be search recommendations based on previous searches and interactions on the platform but could also explicitly match companies' needs to other companies' products with the goal of creating a symbiotic relationship.

Our B2B trading platform could, in the future, be further enhanced by adding matchmaking functionalities. This system could leverage search history and platform interactions to match companies with listings or companies with other companies. We believe there are two possible approaches:

1. The platform could make personalized recommendations on the homepage in a similar way to Youtube or E-Commerce websites.
2. The platform could process a company's defined needs and match them with complementary products or services. This targeted matching could foster the creation of symbiotic relationships, where both parties benefit from the matching.

9.3.2 Reranker

As noted in subsection 8.4.2, we believe that the search accuracy of the semantic search could be improved with a reranker. This approach could also be used for the other two information retrieval systems. A reranker is a machine learning model that takes the retrieved results and re-orders them based on the relevance between the query and the result, as shown in Figure 9.1 below. This could be a large language model (LLM), making it similar to reordering by our "LLM relevancy score" that we used in our automated evaluation, or it could be another transformer model, such as a cross-encoder that has been specifically trained for the task of ranking similarity of texts. This approach could significantly improve the overall effectiveness as the reranker can understand the contextual meaning of the query and the results in its non-condensed form, as the embedding to a vector can have information loss.



Figure 9.1: An overview of how a reranker could be added to the retrieval process. Image source: [55].

9.3.3 Query Transformations

There are many ways to transform and enhance queries [54]. The goal of these methods is to improve the retrieval process by using techniques such as retrieving results using multiple phrasings of the query or breaking down queries into smaller subqueries. For example, the query “An oil filter for the X5 transmission” could be broken down into two queries, “oil filter” and “X5 transmission”. However, most of these techniques require an extra step or a reranker after the retrieval process to combine the results of the transformed queries back into one accurate list of results.

9.3.4 Hybrid Search

As we have already noted in our conclusions, we believe that a hybrid approach leveraging both Elasticsearch and semantic search would be a great solution as they both have their advantages and disadvantages. Elastic search for when queries contain many keywords and semantic search for when queries require more contextual understanding. A hybrid search approach, such as with Reciprocal Rank Fusion (RRF) [56], would combine the strengths of both Elasticsearch and semantic search to deliver a better and more relevant search experience, as illustrated in Figure 9.2 below.

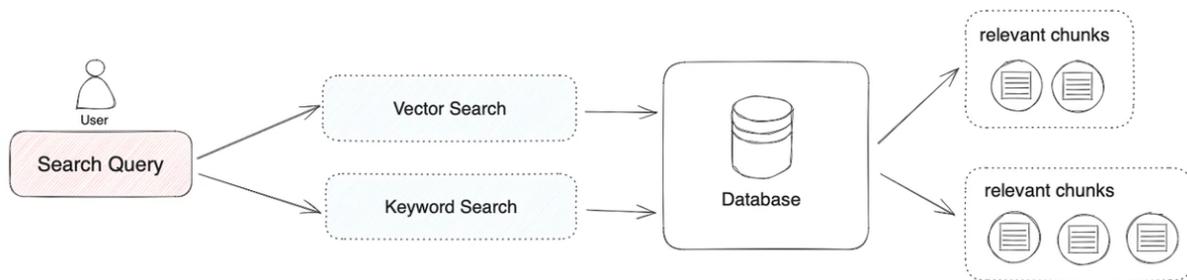


Figure 9.2: Illustration of a hybrid approach combining semantic search and keyword search.
Image source: [57]

9.3.5 Usability & Trustworthiness

Enhancing the user experience and improving the trustworthiness of search results are key areas for future exploration. The following are potential avenues for further research:

- **Result Validation:** Currently, users have limited ways to assess the relevance of search results. Features like keyword highlighting can help users understand why a particular document was returned. This transparency can build trust in the search functionality.
- **Confidence Scoring:** Implementing a confidence score or percentage match can provide users with a clearer understanding of how closely a document aligns with their query. This could also be our “LLM relevance score”, which we used in subsection 8.4.2 as part of our automated evaluation.
- **Query Suggestions:** As users type their search queries, the system could offer suggestions for similar or more precise terms. This can help users refine their searches and potentially discover more relevant results.

These additions can significantly improve the search experience by empowering users to validate results, understand their relevance, and refine their search strategies.

9.3.6 Production

The current state of the project is only meant for development. There are a few steps that have to be taken to have a production-ready platform.

Dockerfiles

The Dockerfiles currently being used to define the Docker images are not optimized for a production environment. These Dockerfiles may not be optimized for the amount of use which could be expected in production. We recommend waiting until the platform goes into production to make any purposeful changes to the Dockerfiles, as the setup and the requirements may change during the next iterations of teams working on the project.

Deployment Pipeline

The current pipeline is a manual deployment, as there was no use case to justify such a time-intensive task of setting up an automated deployment pipeline. We note that the following improvements could be made:

- Create a pipeline that builds the Docker images and deploys them to a container registry with a proper version tag.
- Create a pipeline that deploys the newly generated images if all the checks have run successfully.
- Create a pipeline that notifies the developer if a check fails.
- Create a pipeline that starts a specific deployment for each merge request to be able to have a preview of the changes in a non-local environment.

We extend our sincere gratitude to everyone who supported our work. We are confident that future development and research will unlock the full potential of our B2B trading platform.

Declaration of Honesty

We each hereby declare that we have written the present bachelor thesis independently and only using the specified sources. The passages taken from the listed sources, either literally or in terms of content, are marked in the thesis as a quotation or paraphrase. This bachelor thesis has not yet been published. It has therefore neither been made accessible to other interested parties nor submitted to another examination board.

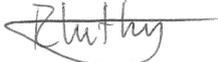
In the spirit of transparency, we would like to acknowledge the use of certain tools to enhance our work. We utilized GitHub Copilot for code generation and suggestion, Grammarly for grammar and style corrections, and Large Language Models like ChatGPT for exploring alternative sentence structures and rephrasing.

Windisch, 22. März 2024

Name: George Rowlands

Unterschrift: 

Name: Raphael Lüthy

Unterschrift: 

References

- [1] India Brand Equity Foundation. *Infosys Becomes 4th Indian Company To Touch \$100 bn Market Cap*. <https://www.ibef.org/news/infosys-becomes-4th-indian-company-to-touch-100-bn-market-cap>. Accessed: 2024-03-16. 2024 (cit. on p. 1).
- [2] Worldmeter. *World Population Clock*. <https://www.worldometers.info/world-population/>. Accessed: 2024-03-16. 2024 (cit. on p. 1).
- [3] Eidgenoessisches Finanzdepartement EFD. *Uebernahme der Credit Suisse durch die UBS*. <https://www.efd.admin.ch/efd/de/home/finanzplatz/uebernahme-credit-suisse-ubs.html>. Accessed: 2024-03-16. 2024 (cit. on p. 1).
- [4] Switzerland Global Enterprise. *Switzerland again leads the world in patents*. <https://www.sge.com/en/article/news/20231-ranking-european-patent-office?ct>. Accessed: 2024-03-16. 2024 (cit. on p. 1).
- [5] Clare Hampel. *globalEDGE Blog: Switzerland's Business Trends: Unique Advantages for International Companies*. [https://globaledge.msu.edu/blog/post/57333/switzerland's-business-trends--unique-advantages-for-international-companies-](https://globaledge.msu.edu/blog/post/57333/switzerland-s-business-trends--unique-advantages-for-international-companies-). Accessed: 2024-03-16. 2024 (cit. on p. 1).
- [6] BBC. *India signs \$100bn free trade deal with four European nations*. <https://www.bbc.com/news/world-asia-68531241>. Accessed: 2024-03-16. 2024 (cit. on p. 1).
- [7] Fortune Business Insights. *Natural Language Processing (NLP) Market Size Growth, 2032*. <https://www.fortunebusinessinsights.com/industry-reports/natural-language-processing-nlp-market-101933>. Accessed: 2024-03-16. 2024 (cit. on p. 1).
- [8] Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. „Attention Is All You Need“. In: (2017). arXiv: 1706.03762 [astro-ph.IM]. URL: <http://arxiv.org/pdf/1706.03762.pdf> (cit. on pp. 1, 16).
- [9] R. Guha, Rob McCool, and Eric Miller. „Semantic search“. In: *Proceedings of the twelfth international conference on World Wide Web - WWW '03*. WWW '03. ACM Press, 2003. DOI: 10.1145/775152.775250. URL: <http://dx.doi.org/10.1145/775152.775250> (cit. on pp. 1, 10).
- [10] Alec Radford. *Improving language understanding with unsupervised learning*. <https://openai.com/research/language-unsupervised>. Accessed: 2024-03-16. 2024 (cit. on pp. 1, 9, 16).
- [11] Richard Elsasser. „Bridging the Gap, An Online Platform for Swiss-Indian Startup Collaboration“. MA thesis. FHNW School of Business, 2023 (cit. on p. 2).
- [12] arvindpdmn. *Information Retrieval*. <https://devopedia.org/information-retrieval>. Accessed: 2024-03-17. 2024 (cit. on p. 4).
- [13] Nick Babich. *Best Practices for Search Results*. <https://uxplanet.org/best-practices-for-search-results-1bbed9d7a311>. Accessed: 2024-03-18. 2024 (cit. on p. 8).
- [14] FannyV. *Search UX Best Practices*. <https://pencilandpaper.io/articles/search-ux/>. Accessed: 2024-03-18. 2024 (cit. on p. 8).
- [15] Tomas Mikolov, Kai Chen, Greg Corrado, et al. „Efficient Estimation of Word Representations in Vector Space“. In: (2013). arXiv: 1301.3781 [astro-ph.IM]. URL: <http://arxiv.org/pdf/1301.3781.pdf> (cit. on p. 9).
- [16] Nils Reimers and Iryna Gurevych. „Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks“. In: (2019). arXiv: 1908.10084 [astro-ph.IM]. URL: <http://arxiv.org/pdf/1908.10084.pdf> (cit. on p. 9).

- [17] Jacob Devlin, Ming-Wei Chang, Kenton Lee, et al. „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding“. In: (2018). arXiv: 1810 . 04805 [astro-ph.IM]. URL: <http://arxiv.org/pdf/1810.04805.pdf> (cit. on p. 9).
- [18] Wenhao Zhu, Hongyi Liu, Qingxiu Dong, et al. „Multilingual Machine Translation with Large Language Models: Empirical Results and Analysis“. In: (2023). arXiv: 2304 . 04675 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2304.04675> (cit. on p. 9).
- [19] Tom B. Brown, Benjamin Mann, Nick Ryder, et al. „Language Models are Few-Shot Learners“. In: (2020). arXiv: 2005 . 14165 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2005.14165.pdf> (cit. on p. 9).
- [20] Luis Serrano. *What Are Word and Sentence Embeddings?* <https://txt.cohere.com/sentence-word-embeddings/>. Accessed: 2024-03-15. 2024 (cit. on p. 9).
- [21] Patrick Lewis, Ethan Perez, Aleksandra Piktus, et al. „Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks“. In: (2020). arXiv: 2005 . 11401 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2005.11401.pdf> (cit. on p. 10).
- [22] Ziwei Xu, Sanjay Jain, and Mohan Kankanhalli. „Hallucination is Inevitable: An Innate Limitation of Large Language Models“. In: (2024). arXiv: 2401 . 11817 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2401.11817> (cit. on p. 10).
- [23] Frederic Agnes Jeremy Ferrero. *Using Word Embedding for Cross-Language Plagiarism Detection*. <https://aclanthology.org/E17-2066.pdf>. Accessed: 2024-03-17. 2024 (cit. on p. 10).
- [24] Xiangyu Zhao, Maolin Wang, Xinjian Zhao, et al. „Embedding in Recommender Systems: A Survey“. In: (2023). arXiv: 2310 . 18608 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2310.18608> (cit. on p. 10).
- [25] Alec Radford, Jong Wook Kim, Chris Hallacy, et al. „Learning Transferable Visual Models From Natural Language Supervision“. In: (2021). arXiv: 2103 . 00020 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2103.00020.pdf> (cit. on p. 10).
- [26] Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, et al. „Text-to-image Diffusion Models in Generative AI: A Survey“. In: (2023). arXiv: 2303 . 07909 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2303.07909> (cit. on p. 10).
- [27] Uriel Singer, Adam Polyak, Thomas Hayes, et al. „Make-A-Video: Text-to-Video Generation without Text-Video Data“. In: (2022). arXiv: 2209 . 14792 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2209.14792.pdf> (cit. on p. 10).
- [28] Elasticsearch Documentation. *What is Semantic Search? A Comprehensive Semantic Search Guide*. <https://www.elastic.co/what-is/semantic-search>. Accessed: 2024-03-19. 2024 (cit. on p. 10).
- [29] Google Research. *Information Retrieval and the Web*. <https://research.google/research-areas/information-retrieval-and-the-web/>. Accessed: 2024-03-17. 2024 (cit. on p. 11).
- [30] Haixun Wang and Taesik Na. „Rethinking E-Commerce Search“. In: (2023). arXiv: 2312 . 03217 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2312.03217> (cit. on p. 11).
- [31] Harald Steck et al. „Deep learning for recommender systems: A Netflix case study“. In: *AI Magazine* 42.3 (Sept. 2021), pp. 7–18. ISSN: 2371-9621. DOI: 10.1609/aimag.v42i3.18140. URL: <http://dx.doi.org/10.1609/aimag.v42i3.18140> (cit. on p. 11).
- [32] Sergey Brin Lawrence Page. *The PageRank Citation Ranking: Bringing Order to the Web*. <https://www.eecs.harvard.edu/~michaelm/CS222/pagerank.pdf>. Accessed: 2024-03-17. 2024 (cit. on p. 11).
- [33] Google Research. *Introducing the Knowledge Graph: things, not strings*. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>. Accessed: 2024-03-17. 2024 (cit. on p. 11).

- [34] Kailash A. Hambarde and Hugo Proença. „Information Retrieval: Recent Advances and Beyond“. In: *IEEE Access* 11 (2023), pp. 76581–76604. ISSN: 2169-3536. DOI: 10.1109/access.2023.3295776. URL: <http://dx.doi.org/10.1109/ACCESS.2023.3295776> (cit. on p. 11).
- [35] Anand Rajaraman and Jeffrey David Ullman. „Mining of Massive Datasets“. In: *Mining of Massive Datasets*. Cambridge University Press, 2011, pp. 1–17. DOI: 10.1017/cbo9781139058452.002. URL: <http://dx.doi.org/10.1017/CBO9781139058452.002> (cit. on p. 11).
- [36] Stephen Robertson and Hugo Zaragoza. *The Probabilistic Relevance Framework: BM25 and Beyond*. https://www.staff.city.ac.uk/~sbrp622/papers/foundations_bm25_review.pdf. Accessed: 2024-03-17. 2024 (cit. on p. 11).
- [37] PostgreSQLGlobalDevelopmentGroup. *PostgreSQL: Documentation: 12 Text Search Types*. Accessed: 2024-01-30. Feb. 2024. URL: <https://www.postgresql.org/docs/16/textsearch.html> (cit. on p. 13).
- [38] Laurent Cazanove. *When does Postgres stop being good enough for full text search?* <https://blog.meilisearch.com/postgres-full-text-search-limitations/>. Accessed: 2024-02-26. 2023 (cit. on p. 13).
- [39] Bo Andersen. *Understanding Sharding in Elasticsearch*. <https://codingexplained.com/coding/elasticsearch/understanding-sharding-in-elasticsearch>. Accessed: 2024-03-03. 2024 (cit. on p. 14).
- [40] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, et al. „ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness“. In: (2018). arXiv: 1811.12231 [astro-ph.IM]. URL: <http://arxiv.org/pdf/1811.12231> (cit. on p. 15).
- [41] Hadas Kotek. *Gender bias and stereotypes in Large Language Models*. <https://dl.acm.org/doi/fullHtml/10.1145/3582269.3615599>. Accessed: 2024-03-17. 2024 (cit. on p. 15).
- [42] Ondra Urban. *Is web scraping legal? Yes, if you know the rules*. <https://blog.apify.com/is-web-scraping-legal/#is-scraping-copyrighted-content-legal>. Accessed: 2024-03-03. 2024 (cit. on p. 16).
- [43] Sondos Mahmoud Bsharat, Aidar Myrzakhan, and Zhiqiang Shen. „Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4“. In: (2023). arXiv: 2312.16171 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2312.16171> (cit. on p. 18).
- [44] Jason Wei, Xuezhi Wang, Dale Schuurmans, et al. „Chain-of-Thought Prompting Elicits Reasoning in Large Language Models“. In: (2022). arXiv: 2201.11903 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2201.11903> (cit. on p. 18).
- [45] Leland McInnes, John Healy, and James Melville. „UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction“. In: (2018). arXiv: 1802.03426 [astro-ph.IM]. URL: <http://arxiv.org/pdf/1802.03426> (cit. on p. 19).
- [46] Atlassian. *Microservices vs. monolithic architecture*. <https://www.atlassian.com/microservices/microservices-architecture/microservices-vs-monolith>. Accessed: 2024-03-10. 2024 (cit. on p. 23).
- [47] Avi Networks. *What is Auto Scaling? Definition FAQs*. <https://avinetworks.com/glossary/auto-scaling/>. Accessed: 2024-02-27. 2024 (cit. on p. 24).
- [48] Erich Gamma et al. *Design Patterns*. en. Addison-Wesley Professional, Janu. ISBN: 0-201-63361-2 (cit. on p. 33).
- [49] Jake Wells. *Find anything blazingly fast with Google’s vector search technology*. <https://cloud.google.com/blog/topics/developers-practitioners/find-anything-blazingly-fast-googles-vector-search-technology> author=Jake Wells, 2021 (cit. on p. 35).

- [50] Faisal Rahutomo, Teruaki Kitasuka, and Masayoshi Aritsugi. „Semantic Cosine Similarity“. In: Oct. 2012 (cit. on p. 37).
- [51] Christopher Winslett. *HNSW Indexes with Postgres and pgvector*. <https://www.crunchydata.com/blog/hnsw-indexes-with-postgres-and-pgvector>. Accessed: 2024-03-03. 2024 (cit. on p. 38).
- [52] Apple Developer Documentation. *About the Rosetta Translation Environment*. <https://developer.apple.com/documentation/apple-silicon/about-the-rosetta-translation-environment>. Accessed: 2024-03-22. 2024 (cit. on p. 45).
- [53] Omar Khattab and Matei Zaharia. „ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT“. In: (2020). arXiv: 2004.12832 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2004.12832> (cit. on p. 65).
- [54] Langchain Blog. *Query Transformations*. <https://blog.langchain.dev/query-transformations/>. Accessed: 2024-03-18. 2024 (cit. on pp. 65, 70).
- [55] LLamaIndex Blog. *Using LLM's for Retrieval and Reranking*. <https://www.llamaindex.ai/blog/using-llms-for-retrieval-and-reranking-23cf2d3a14b6>. Accessed: 2024-03-19. 2024 (cit. on p. 69).
- [56] Priyanka Mandikal and Raymond Mooney. „Sparse Meets Dense: A Hybrid Approach to Enhance Scientific Document Retrieval“. In: (2024). arXiv: 2401.04055 [astro-ph.IM]. URL: <http://arxiv.org/pdf/2401.04055> (cit. on p. 70).
- [57] Dify Blog. *Introducing Hybrid Search and Rerank to Improve the Retrieval Accuracy of the RAG System*. <https://dify.ai/blog/hybrid-search-rerank-rag-improvement>. Accessed: 2024-03-19. 2024 (cit. on p. 70).
- [58] OpenAI. *ChatGPT*. <https://openai.com/chatgpt>. Accessed: 2024-03-22. 2024 (cit. on p. 77).
- [59] Google DeepMind. *Gemini*. <https://deepmind.google/technologies/gemini/#introduction>. Accessed: 2024-03-22. 2024 (cit. on p. 77).
- [60] Grammarly Inc. *Grammarly: Free AI Writing Assistance*. <https://www.grammarly.com/>. Accessed: 2024-03-22. 2024 (cit. on p. 77).
- [61] Inc. Perplexity AI. *Perplexity AI*. <https://www.perplexity.ai/>. Accessed: 2024-03-22. 2024 (cit. on p. 77).
- [62] GitHub. *GitHub Copilot - Your AI pair programmer*. <https://github.com/features/copilot>. Accessed: 2024-03-22. 2024 (cit. on p. 77).
- [63] OpenAI. *DALL-E 3*. <https://openai.com/dall-e-3>. Accessed: 2024-03-21. 2024 (cit. on p. 77).

List of Tools used

Tool	Usage	Sections
ChatGPT [58]	Used for exploring alternative sentence structures and rephrasing.	All Sections
Google Gemini [59]	Used for exploring alternative sentence structures and rephrasing.	All Sections
Grammarly [60]	Used for grammar and style corrections	All Sections
Perplexity [61]	Used for research and inspiration.	All Sections
Github Copilot [62]	Used for code generation.	Used during the development of the whole platform.
Dall-E 3 [63]	Generation of the cover image.	Titlepage

Table 9.1: Table listing all AI-based tools used during this project.

List of Figures

2.1	An overview of how information retrieval systems work. Image source: [12]	4
3.1	Example of the contextual search check on https://www.alibaba.com/ , accessed: 19.03.2024	6
3.2	Example of a search on https://www.crunchbase.com/ with the results of type “company”, accessed: 19.03.2024	7
3.3	Example of search suggestions and auto-completion on https://www.clutch.co/ , accessed: 19.03.2024	7
3.4	Example of an option to change the language on https://www.wlw.de/ , accessed: 19.03.2024	8
3.5	Visualization of text embeddings from a sentence transformer in a vector space. Image source: [20]	9
3.6	Visualization of how semantic search works to find relevant results. Image source: [28] .	10
3.7	Where q is the query term, d is the document, D is all documents in the corpus, $f(q, d)$ returns the number of times that the term q occurs in document d and $n(q, D)$ returns the number of documents that the term q occurs in.	11
3.8	Where q is the query term, d is the document, D is all documents in the corpus, $avgdl$ the average document length in terms, $f(q, d)$ returns the number of times that the term q occurs in document d and $n(q, D)$ returns the number of documents that the term q occurs in. b and k_1 are free parameters, usually chosen as $b = 0.75$ and $k_1 \in [1.2, 2.0]$. .	12
4.1	Visualization of how Elasticsearch creates an index and splits it into shards. Image source: [39]	14
5.1	A simplified diagram of the prompting and generation process for a company (blue), the products and services it offers (red) and the needs and desires it seeks (green).	18
5.2	Visualization of the text embeddings after being reduced to two dimensions.	19
6.1	Our data models.	20
6.2	A visualization of the difference between a monolith and microservices. Image source: [46]	23
6.3	How auto-scaling works in a simplified way. Image source: [47]	24
6.4	Layering of our microservices.	25
6.5	Wireframe for the search page.	26
7.1	The user interface for the search page. At the top are the category filters; at the bottom, the search bar; on the left are the search results and in the center are the details of the selected search result. The globe in the top right corner is used to set the language. . . .	29
7.2	Generated Swagger Documentation.	31
7.3	Diagram of the interface architecture for the part of the company and listing management service (CLMS).	32
7.4	Visualization of distance functions between vectors. Image source: [49]	35
7.5	Diagram of layered graph traversal. Image source: [51].	38

7.6	Basic Architecture of the sync between the Postgres database and Elasticsearch.	40
7.7	Overview of the End-To-End Testing setup.	43
8.1	Screenshot of the Rosetta emulation layer option in Docker Desktop.	45
8.2	Latency test of our baseline, PostgreSQL FTS.	47
8.3	Latency test of the semantic search with pgvector and the HSNW index.	48
8.4	Latency test of the semantic search with pgvector and without the HSNW index.	49
8.5	Latency Evaluation for Elasticsearch.	50
8.6	An example of an A/B test for the query “cancer”. The semantic search returned the five results at the top. PostgreSQL FTS returned the five results at the bottom.	53
8.7	Breakdown of the survey participants.	54
8.8	The different business sectors of all survey participants.	55
8.9	The total vote distribution for all three information retrieval systems and voting schemes.	56
8.10	Analysis of whether participants found our search results for the query “a phone with 256 GB storage” better than other search tools.	57
8.11	The distribution of where the expected result is found.	61
8.12	The hit rate gain when more search results are considered.	61
8.13	Violin plots showing the distribution and median of the LLM relevance score for the top five search results of each of our search approaches, <i>i.e.</i> the top-ranked search results. In the centre of each violin is a boxplot, with the black dot being the median value of the data and the box forming the interquartile range of the data. The data range and distribution can also be interpreted from the violin’s shape (a histogram on its side). A wide area showing where data points are more frequent.	63
8.14	A histogram of the standard deviation values of all LLM relevancy scores.	64
8.15	The cosine distance of search results to the search query is very tightly packed together, making the choice difficult.	66
9.1	An overview of how a reranker could be added to the retrieval process. Image source: [55].	69
9.2	Illustration of a hybrid approach combining semantic search and keyword search. Image source: [57]	70
B.1	The wireframe for the login page.	84
B.2	The wireframe for the registration page.	84
B.3	The wireframe for the email verification page.	85
B.4	The wireframe for the page to join an existing business.	85
B.5	The wireframe for the page to create a business.	86

List of Tables

- 3.1 Feature comparison of five of the current B2B trading platforms. 8
- 4.1 Comparison of our three selected information retrieval systems. ES is the abbreviation for Elasticsearch. 13
- 8.1 Latency comparison of our three search approaches. SS is the short form for Semantic Search. ES is the short form for Elasticsearch. 51
- 8.2 The amount of votes that the bold search approach above lost to the search approach on the left. 58
- 8.3 Results of the automated evaluation using all 80 queries for each of the three search approaches. 60
- 8.4 Results of the automated evaluation split between the human-written queries and the queries the LLM generated. 62
- 9.1 Table listing all AI-based tools used during this project. 77

List of Listings

1	Example of a business sector and its description.	16
2	Generating structured output with the Instructor package and parsing it to a valid Pydantic model.	17
3	Start of the english.json translation file.	30
4	Example of the translation hook used for the login page title.	30
5	Extract of an interface for listing service	32
6	Initialization of the specific IRS implementation.	33
7	The code for searching for companies. The name_tokens and description_tokens are the ts_vectors.	34
8	The EmbeddingService without the communication layer.	36
9	The interface defining the search parameters for the listing service in the company and listing management service (CLMS).	37
10	Database query for the vector nearest neighbour search using pgvector.	37
11	How to configure pgvector to use the HNSW index.	39
12	Example of how the name and description of a listing are merged before creating the text embedding.	39
13	PGSync schema for synchronizing PostgreSQL to Elasticsearch.	41
14	Example of a test case to test the creation of a company. The test data is anonymised for publication purposes.	44
15	Shortened version of the latency testing script	46
16	CSV output example	46
17	Prompts we used to generate a company.	82
18	Prompts we used to generate a company's products.	82
19	Prompts we used to generate a company's needs.	83
20	The prompt used for the LLM relevancy score.	83

A Prompts Appendix

```

# Business sector context
sector_system_prompt = f"""
When talking about the {current_sector} sector then I mean companies
fulfill the following description:
{sector_description}
"""

# Generate company name and description
list_of_company_names_prompt = f"""
Generate a list of company names that are unique and creative. The
companies work in {current_sector} sector. Do not use any names that
have the word "innovative" or "creative"
in them.
"""

company_name_description_prompt = f"""
Pick one of these company names and return it along with a description
of the company that is very detailed and extensive in JSON. The company
description should be between 2500 and 3000 characters long and contain
something about the company's history, what they do, and what they are
known for and their sustainability efforts. Write the description that
is either neutral or from the company's perspective.
"""

```

Listing 17: Prompts we used to generate a company.

```

# Generate products and services of the company
product_ideas = f"""
Given the following company name and description, generate a possible
product or service that the company could offer.
{company.name}
{company.description}
"""

product_enhancement = f"""
Enhance the product description by adding more details to it.
Make the description more elaborate and versatile. Cover technical
specifications. Price range, customization options, delivery options,
waiting time, etc. Be creative in thinking about any missing
product or service attributes.
"""

```

Listing 18: Prompts we used to generate a company's products.

```

# Generate needs of the company
need_ideas = f"""
Given the following company name and description, generate a need
the company could have from an other company. Be creative in
thinking about any gaps the company might have. A need has to have
a name and description.
{company.name}
{company.description}
"""

need_enhancement = f"""
Enhance the description by adding more details to it.
Make the description more elaborate and versatile. Cover technical
specifications. Price range, customization options.
"""

```

Listing 19: Prompts we used to generate a company's needs.

```

prompt = f"""
You are a RELEVANCE grader; providing the relevance of the given
RESULT to the given QUERY. Respond only as a number from 0 to 10
where 0 is the least relevant and 10 is the most relevant.
A few additional scoring guidelines:
- Long RESULTS should score equally well as short RESULTS.
- RELEVANCE score should increase as the RESULT provides more
RELEVANT context to the QUERY.
- RELEVANCE score should increase as the RESULT provides RELEVANT
context to more parts of the QUERY.
- RESULT that is RELEVANT to some of the QUERY should score of 2,
3 or 4. Higher score indicates more RELEVANCE.
- RESULT that is RELEVANT to most of the QUERY should get a score
of 5, 6, 7 or 8. Higher score indicates more RELEVANCE.
- RESULT that is RELEVANT to the entire QUERY should get a score
of 9 or 10. Higher score indicates more RELEVANCE.
- RESULT must be relevant and helpful for answering the entire
QUERY to get a score of 10.
- Answers that intentionally do not answer the question, such as
'I don't know', should also be counted as the most relevant.
- Never elaborate.

QUERY: {query}
RESULT: {name} - {description}
RELEVANCE:
"""

```

Listing 20: The prompt used for the LLM relevancy score.

B Onboarding Wireframes Appendix

Login

The wireframe for the login page is contained within a rounded rectangular border. It features the following elements:

- SMEConnect**: The brand name in a bold, sans-serif font.
- Log in to your account**: A sub-header in a smaller font.
- Email**: A text input field.
- Password**: A text input field.
- Remember me**: A checkbox followed by the text "Remember me".
- [Forget password?](#): A blue link for password recovery.
- Log in**: A button for submitting the login form.
- Don't have an account? [Create an account](#)**: A link for new users, with "Create an account" in blue.

Figure B.1: The wireframe for the login page.

RegisterAccount

The wireframe for the registration page is contained within a rounded rectangular border. It features the following elements:

- Registration**: The title of the form in a bold, sans-serif font.
- Firstname** and **Lastname**: Two side-by-side text input fields.
- Email**: A text input field.
- Password**: A text input field with a small square checkbox to its right.
- Confirm Password**: A text input field with a small square checkbox to its right.
- Register**: A button for submitting the registration form.

Figure B.2: The wireframe for the registration page.

VerifyEmail

<p>Verify your email address</p> <p>Please click on the link in the email we just sent you to confirm your email address.</p> <p>Resend email</p>	
---	--

Figure B.3: The wireframe for the email verification page.

JoinBusiness

<p>Join or create a business</p> <p>Does your business already exist on SMEConnect? Then add the business id below to request to join.</p> <p><input type="text" value="Business ID"/> ?</p> <p><input type="button" value="Join"/></p> <p>Are you the business owner and want to create a new business on SMEConnect?</p> <p>Create business</p>	
---	--

Figure B.4: The wireframe for the page to join an existing business.

CreateBusiness

Create a new business

Logo or image ?

Company name

City, Country

Business Sector

Subsector?

Create

Figure B.5: The wireframe for the page to create a business.

C Survey Appendix

Bridging the Gap

Comparing Text Retrieval Systems for Companies & Products

Dear Participant,

Thank you sincerely for dedicating your time to assist us in the evaluation of our bachelor thesis project.

To provide you with a brief overview, our bachelor thesis revolves around the development of a platform designed to facilitate global business connections. **This platform serves as a place for companies worldwide to engage with one another, enabling them to identify compatible partners, products, and services tailored to their specific needs.** In our thesis, **we focus on the platform's search and exploration capabilities** which is essential to the success of this platform. So far most platforms for Business to Business (B2B) trading have very inaccurate searches, which is what we want to improve on.

Your valuable insights and feedback are instrumental in ensuring that the platform's search and exploration capabilities can be as accurate as possible and provide relevant results to the users.

The survey should not take longer than 15 minutes to complete and can be done no matter the role! Our focus is on A/B testing, where **you are given two different search results and should choose which set of search results you find more accurate or relevant to the given search query.**

IMPORTANT: To better read the search results you can right-click on the image of the search results and select "Open Image in new tab", there you will see the full-size image.

* 1. Name

* First name

* Last name

* 2. Business Sector

- Information Technology (IT) and Software Development
- Engineering
- Manufacturing
- Energy and Utilities
- Financial Services
- Healthcare and Pharmaceuticals
- Other (please specify)

* 3. Are you currently or have you previously been directly involved in Business to Business (B2B) trading?

- Yes

No

1 / 8

12%

Next

Powered by



See how easy it is to [create surveys and forms](#).

[Privacy & Cookie Notice](#)

Bridging the Gap

Comparing Search Results

In the following questions, we will do a **comparison of our searching solutions**.

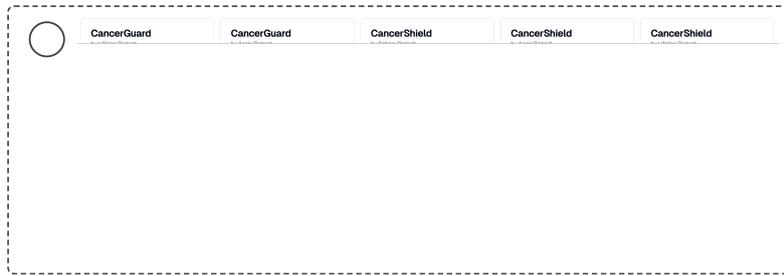
You will see two images where each picture represents search results from a different searching solution. In a picture, you will see 5 products or services our search solutions deem to be the most relevant to the query. **Please compare the results and select, which one of the two you find more relevant for the search query that was given.** For each search query we will provide a possible use case at the beginning of the section.

You will see search results multiple times, since we have three approaches we want to compare with each other. **You will also see some images that contain no results and the following text: "Sorry, we couldn't find any results for your search :("**.

The search queries will differ in complexity. This is to accommodate the different types of searches users may use.

The first search results we will compare are for the search query "cancer". You could imagine this query being used if for example a laboratory needs equipment to research cancer and hopefully find treatments.

* 4. Results 1 of search: "Cancer"



<p>GenoCure By BioFourish</p> <p>GenoCure is a personalized treatment plan based on a patient's genetic profile. Our team of experts analyze the patient's genome to identify genetic variations that might contribute to the disease. Once identified, a personalized treatment plan is developed to target these variations. GenoCure can be used for a variety of diseases, including cancer, diabetes, and cardiovascular diseases.</p>	<p>BioLuminate Oncology... By BioLuminate</p> <p>Our Oncology Therapeutics range is a revolution in cancer treatment. Utilizing breakthroughs in genetic and molecular science, we offer targeted therapy medicines that work by interfering with specific proteins, genes, or processes that contribute to the growth and survival of cancer cells. Our innovative range includes solutions for a variety of cancer types including breast, lung, and colorectal cancers. These...</p>	<p>LuminoTherapy By BioLuminate</p> <p>LuminoTherapy is a state-of-the-art cancer treatment that utilizes the power of light to target and eliminate cancer cells. This non-invasive procedure has been proven to be highly effective in curing various types of cancers, with minimal side effects. The therapy involves injecting patients with a photosensitive drug, followed by targeted light radiation. LuminoTherapy has been extensively tested and...</p>	<p>LuminoTherapy By BioLuminate</p> <p>LuminoTherapy is BioLuminate's flagship product, a revolutionary new form of therapy for various types of cancer. Utilizing our patented, cutting-edge light technology, LuminoTherapy targets and destroys cancer cells with remarkable precision, minimizing the damage to healthy cells. It's a non-invasive and patient-friendly alternative to traditional chemotherapy and radiation. The therapy can b...</p>	<p>GenomeShield By Vitality Biosciences</p> <p>GenomeShield is a cutting-edge genetic editing technology offered by Vitality Biosciences. It is designed to provide highly precise and efficient genetic modifications that can aid in the treatment of a wide array of illnesses. GenomeShield leverages advanced CRISPR-Cas9 technology to perform exact genetic alterations, opening the door to targeted therapies for conditions like cancer, neurodegenerative...</p>
--	--	--	--	--

Both search results feel equal

* 5. Results 2 of search: "Cancer"



<p>BioTherapy X By BioNexus</p> <p>BioTherapy X is a groundbreaking oncology drug developed by BioNexus. This novel therapeutic is designed to target and kill cancer cells with unprecedented precision, reducing the side effects commonly associated with traditional cancer treatments. BioTherapy X represents a new era in cancer treatment, providing hope to patients with hard-to-treat forms of cancer.</p>	<p>BioDose Immunotherapy By BioDose</p> <p>BioDose is an advanced immunotherapy treatment option for cancer patients. This personalized therapy uses the body's immune system to fight cancer cells effectively. It involves the generation of patient-specific vaccines using the patient's cancer cells and immune cells. The vaccines are then reintroduced into the patient's body to stimulate an immune response against the cancer cells. BioDose is a...</p>	<p>OncoSphere By Sphere Biotech</p> <p>OncoSphere is a revolutionary cancer therapy developed by Sphere Biotech. This innovative treatment utilizes advanced biotechnology to target and destroy cancer cells, while sparing healthy cells. It is designed to treat a variety of cancers, including breast, lung, and colon cancer. OncoSphere is a testament to Sphere Biotech's commitment to improving human health and its...</p>	<p>Oncoscan By NeoScan</p> <p>Oncoscan is a comprehensive diagnostic tool designed to identify and monitor various forms of cancer. Using advanced biotechnology, Oncoscan can detect even the smallest traces of cancer cells in the body, providing the most accurate and early diagnosis possible. Oncoscan's cancer-detection capabilities can identify a wide range of cancer types, including lung, breast, colon, and prostate cancers. It's no...</p>	<p>BioShield Cancer... By VitaShield</p> <p>BioShield Cancer Therapy is a revolutionary treatment developed by VitaShield Biotech. This therapy uses our patented biotechnology to target and eliminate cancer cells with precision, thereby reducing the side effects commonly associated with traditional cancer treatments. The therapy is customizable to each patient's unique genetic makeup and cancer type, ensuring a personalized approach to treatment. The...</p>
--	---	--	--	--

Both search results feel equal

* 6. Results 3 of search: "Cancer"

BioTherapy X
By BioRexus

BioDose Immunotherapy
By Pulse BioSolutions

OncoSphere
By Sphere Biotech

Oncoscan
By Aegis Biotech

BioShield Cancer...
By Lifeline Biotech

CancerGuard
By Lifeline Biotech

CancerGuard
By Lifeline Biotech

CancerShield
By Sphere Biotech

CancerShield
By Aegis Biotech

CancerShield
By Lifeline Biotech

Both search results feel equal

* 7. These search results are better than other search tools I have used.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

2 / 8

25%

Prev

Next

Powered by



See how easy it is to [create surveys and forms](#).

Bridging the Gap

Results Comparison of "A phone with 256 GB storage"

This search query has more natural language than the previous one for "cancer". The search query could be thought of as a query an employer might use if he wants to bulk phones for his/her employees.

* 8. Results 1 of search: "A phone with 256 GB storage"

<input type="radio"/> <p>Eco-Tech Smartphone <small>by Product Reviews</small></p> <p>The Eco-Tech Smartphone is a cutting-edge device that merges technology and sustainability. It features a 6.5-inch OLED display, a powerful quad-core processor, and a long-lasting battery. The smartphone is designed with recycled materials, making it an environmentally-friendly choice. The device is customizable with various color options and storage capacities ranging from 64G...</p>	<input type="radio"/> <p>Solaris Smartphone <small>by Device Reviews</small></p> <p>The Solaris Smartphone is the latest addition to our electronics line. Featuring a 6.5-inch Super AMOLED display and powered by a cutting-edge octa-core processor, this device is designed for speed and performance. It comes with a 48MP primary camera, an ultra-wide-angle lens, and a depth sensor for capturing stunning photos and videos. It is pre-installed with the latest Android version and...</p>	<input type="radio"/> <p>Nexus Cloud <small>by Server Clouds</small></p> <p>Nexus Cloud is our advanced cloud computing service. It is designed to provide businesses with scalable computing power, robust storage, and advanced networking capabilities. With Nexus Cloud, businesses can host their applications in a secure, high-performing, and reliable environment. The service offers a range of features, including virtual private servers, database services, and content deliv...</p>	<input type="radio"/> <p>Nexus Cloud <small>by Server Clouds</small></p> <p>Nexus Cloud is our comprehensive cloud computing solution designed to help businesses scale their operations with ease. It offers a range of services including Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). With Nexus Cloud, businesses have the flexibility to choose the services that best fit their needs and only pay for what they use. Our cloud services...</p>	<input type="radio"/> <p>FusionMobile <small>by CloudDev</small></p> <p>FusionMobile is a mobile application development service that helps businesses to create innovative, user-friendly, and secure mobile applications. Our team of experts uses the latest technologies and best practices to develop applications that offer a seamless user experience across different platforms such as iOS, Android, and Windows. FusionMobile offers various customization...</p>
<input type="radio"/> <p>Beacon Wealth... <small>by Capital Beacon</small></p> <p>Our Beacon Wealth Management is a comprehensive financial planning service tailored to meet the unique needs of high-net-worth individuals and corporations. It includes desktop management, tax planning, estate planning, and risk management. Our expert wealth managers employ a mix of traditional and alternative investment strategies to optimize returns and minimize risks. The...</p>	<input type="radio"/> <p>Nexus Wealth... <small>by Equity Nexus</small></p> <p>Our Nexus Wealth Management service is a comprehensive financial planning service designed to help clients achieve their financial goals. Our skilled and experienced financial advisors work closely with clients to understand their financial situation, goals, and risk tolerance, and then create a personalized investment strategy that aligns with these factors. This service includes asset...</p>	<input type="radio"/> <p>Wealthy Horizons <small>by Monetary Matrix</small></p> <p>Wealthy Horizons is a premium wealth management service offered by Monetary Matrix. This service is designed for high net worth individuals and businesses who seek to optimize their wealth and financial resources. Wealthy Horizons provides clients with a dedicated wealth manager who will work closely with them to understand their financial goals, risk tolerance, and investment preferences...</p>	<input type="radio"/> <p>Quantum Cloud <small>by Quantum Storage</small></p> <p>Quantum Cloud is a state-of-the-art cloud storage solution designed to provide businesses with secure, scalable, and efficient data storage capabilities. With Quantum Cloud, businesses can store and access their data anytime, anywhere, and from any device. The service features end-to-end encryption, ensuring the highest level of security for your data. Quantum Cloud also offers flexible storage...</p>	<input type="radio"/> <p>SpectraCloud <small>by SpectraCode</small></p> <p>SpectraCloud is a comprehensive cloud service offered by SpectraCode. This service provides businesses with secure and scalable cloud storage solutions, backup services, and disaster recovery options. SpectraCloud is designed to meet the needs of businesses of all sizes, from small start-ups to large corporations. It offers a range of storage options, from basic packages of 100GB to custom...</p>

Both search results feel equal

* 9. Results 2 of search: "A phone with 256 GB storage"



Sorry, we couldn't find any results for your search :(



Beacon Wealth.. <small>by Capital Beacon</small> Our Beacon Wealth Management is a comprehensive financial planning service tailored to meet the unique needs of high-net-worth individuals and corporations. It includes portfolio management, tax planning, estate planning, and risk management. Our expert wealth managers employ a mix of traditional and alternative investment strategies to optimize returns and minimize risks. The...	Nexus Wealth.. <small>by Equity Nexus</small> Our Nexus Wealth Management service is a comprehensive financial planning service designed to help clients achieve their financial goals. Our skilled and experienced financial advisors work closely with clients to understand their financial situation, goals, and risk tolerance, and then create a personalized investment strategy that aligns with these factors. This service includes asset...	Wealthy Horizons <small>by Monetary Matrix</small> Wealthy Horizons is a premium wealth management service offered by Monetary Matrix. This service is designed for high net worth individuals and businesses who seek to optimize their wealth and financial resources. Wealthy Horizons provides clients with a dedicated wealth manager who will work closely with them to understand their financial goals, risk tolerance, and investment preferences...	Quantum Cloud <small>by Quantum Storage</small> Quantum Cloud is a state-of-the-art cloud storage solution designed to provide businesses with secure, scalable, and efficient data storage capabilities. With Quantum Cloud, businesses can store and access their data anytime, anywhere, and from any device. The service features end-to-end encryption, ensuring the highest level of security for your data. Quantum Cloud also offers flexible storage...	SpectraCloud <small>by SpectraCode</small> SpectraCloud is a comprehensive cloud service offered by SpectraCode. This service provides businesses with secure and scalable cloud storage solutions, backup services, and disaster recovery options. SpectraCloud is designed to meet the needs of businesses of all sizes, from small start-ups to large corporations. It offers a range of storage options, from basic packages of 100GB to custom...
--	---	--	---	---



Both search results feel equal

* 10. Results 3 of search: "A phone with 256 GB storage"



Eco-Tech Smartphone <small>by Prodigal Production</small> The Eco-Tech Smartphone is a cutting-edge device that merges technology and sustainability. It features a 6.5-inch OLED display, a powerful quad-core processor, and a long-lasting battery. The smartphone is designed with recycled materials, making it an environmentally-friendly choice. The device is customizable with various color options and storage capacities ranging from 64G...	Solaris Smartphone <small>by Catalyst Productions</small> The Solaris Smartphone is the latest addition to our electronics line. Featuring a 6.6-inch Super AMOLED display and powered by a cutting-edge octa-core processor, this device is designed for speed and performance. It comes with a 48MP primary camera, an ultra-wide-angle lens, and a depth sensor for capturing stunning photos and videos. It is pre-installed with the latest Android version and...	Nexus Cloud <small>by Nexus Codes</small> Nexus Cloud is our advanced cloud computing service. It is designed to provide businesses with scalable computing power, robust storage, and advanced networking capabilities. With Nexus Cloud, businesses can host their applications in a secure, high-performing, and reliable environment. The service offers a range of features, including virtual private servers, database services, and content delive...	Nexus Cloud <small>by Nexus Codes</small> Nexus Cloud is our comprehensive cloud computing solution designed to help businesses scale their operations with ease. It offers a range of services including Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). With Nexus Cloud, businesses have the flexibility to choose the services that best fit their needs and only pay for what they use. Our cloud services...	FusionMobile <small>by FusionByte</small> FusionMobile is a mobile application development service that helps businesses to create innovative, user-friendly, and secure mobile applications. Our team of experts uses the latest technologies and best practices to develop applications that offer a seamless user experience across different platforms such as iOS, Android, and Windows. FusionMobile offers various customization...
--	--	--	---	---



Sorry, we couldn't find any results for your search :(



Both search results feel equal

* 11. These search results are better than other search tools I have used.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

3 / 8

38%

Prev

Next

Powered by



See how easy it is to [create surveys and forms](#).

Bridging the Gap

Results Comparison of "Machine Learning"

Machine learning is being introduced in many fields now. What if a business was looking for a machine learning solution to integrate in their own products?

* 12. Results 1 of search: "Machine learning"

<p>DynaFlex Textile... by Fusion Dynamics</p> <p>The DynaFlex Textile Machine, a flagship product of Fusion Dynamics, is a versatile and high-performance textile manufacturing machine. Designed to handle diverse textile materials, this machine offers unparalleled efficiency and adaptability. It boasts advanced features such as automatic thread count adjustment and fabric tension control, ensuring optimal product quality at all...</p>	<p>OrbitSecure by TechNexus</p> <p>OrbitSecure is a comprehensive cybersecurity solution designed to protect businesses from a wide range of cyber threats. It offers real-time threat detection, security incident management, and rapid response capabilities. With advanced machine learning algorithms, OrbitSecure can identify and neutralize threats before they can cause damage. This service is customizable according to...</p>	<p>CloudGuard Pro by TechNova</p> <p>CloudGuard Pro is an advanced and robust cybersecurity service from TechNova that safeguards your data in the cloud. It utilizes AI and machine learning to monitor and predict potential threats, offering proactive protection against cyber-attacks. CloudGuard Pro can be customized to cater to specific industry requirements and threats, making it an ideal solution f...</p>	<p>Cloud Optimizer by TechTonic</p> <p>Cloud Optimizer is a comprehensive cloud computing solution offered by TechTonic. It allows businesses to manage and optimize their cloud resources effectively. The solution includes features like automatic scaling, resource monitoring, cost management, and security compliance. The technical specifications of the product include compatibility with major cloud platforms...</p>	<p>AI-Boost by TechTonic</p> <p>AI-Boost is a cutting-edge artificial intelligence solution designed to optimize business operations. Powered by advanced machine learning algorithms, AI-Boost can analyze vast amounts of data to provide insightful predictions and recommendations. This enables businesses to make data-driven decisions, automate repetitive tasks, and improve overall efficiency. The solution can...</p>
<p>AI-Boost by TechTonic</p> <p>AI-Boost is a cutting-edge artificial intelligence solution designed to optimize business operations. Powered by advanced machine learning algorithms, AI-Boost can analyze vast amounts of data to provide insightful predictions and recommendations. This enables businesses to make data-driven decisions, automate repetitive tasks, and improve overall efficiency. The solution can...</p>	<p>GeneCode by Genetix Dynamics</p> <p>GeneCode is a complete genomic analysis service that leverages advanced biotechnology to provide a comprehensive understanding of genetic material. From detailed DNA sequencing to in-depth gene-expression profiling, GeneCode delivers precise, actionable insights for researchers and healthcare providers. The service is able to decode the entire human genome in less than 24 hour...</p>	<p>Data-Driven Decision... by Data Drive Systems</p> <p>The Data-Driven Decision Suite is a comprehensive package of software solutions designed to empower businesses with actionable insights. The suite includes a range of tools for data collection, processing, analysis, and visualization. It leverages advanced machine learning algorithms to provide predictive analytics, enabling businesses to anticipate market trends and customer behavior. The suite also...</p>	<p>ProduceTech AutoBots by ProduceTech</p> <p>ProduceTech AutoBots are advanced industrial robots designed for use in automobile manufacturing. They are equipped with state-of-the-art technology to ensure precision and efficiency in assembly line operations. With a wide range of motion and high-speed capabilities, these robots can perform complex tasks such as welding, painting, assembly, pick and place, and quality inspection. They are...</p>	<p>DataSphere by Data Drive Dynamics</p> <p>DataSphere is a comprehensive data management solution designed to simplify and streamline the process of collecting, processing, analyzing, and storing large volumes of data. It offers an intuitive user interface, enabling users to easily navigate through complex data landscapes. With its robust algorithms, DataSphere can quickly process and analyze data...</p>

Both search results feel equal

* 13. Results 2 of search: "Machine learning"

<p>ByteShield by ByteCortex</p> <p>ByteShield is a state-of-the-art cybersecurity solution developed by ByteCortex. It provides top-tier protection against a wide range of threats including malware, ransomware, phishing, and DDoS attacks. With its advanced machine learning algorithms, ByteShield can detect and neutralize threats before they cause harm. It comes with an easy-to-use interface, making it accessible to both IT...</p>	<p>CloudGuard Pro by TechNova</p> <p>CloudGuard Pro is an advanced and robust cybersecurity service from TechNova that safeguards your data in the cloud. It utilizes AI and machine learning to monitor and predict potential threats, offering proactive protection against cyber-attacks. CloudGuard Pro can be customized to cater to specific industry requirements and threats, making it an ideal solution f...</p>	<p>CyberShield by TechNova Dynamics</p> <p>CyberShield is a cutting-edge cybersecurity solution designed to protect businesses from a wide range of digital threats. This robust software suite includes antivirus, anti-malware, firewall, intrusion detection, and data encryption features. It uses advanced machine learning algorithms to detect and neutralize threats in real time, providing comprehensive protection for your digital assets...</p>	<p>CyberShield by TechNova</p> <p>CyberShield is an advanced cybersecurity solution developed by TechNova Dynamics. This state-of-the-art software provides end-to-end protection for your business infrastructure from a wide range of cyber threats. CyberShield is equipped with machine learning algorithms that allow it to detect and neutralize threats in real time. Its user-friendly interface makes it easy to manage and monitor your organization's...</p>	<p>CloudNimbus by TechNimbus</p> <p>CloudNimbus is a robust and scalable cloud computing service offered by TechNimbus. This state-of-the-art service allows businesses to store, manage, and process data on a secure, remote server network. CloudNimbus is built on high-performance computing systems, ensuring fast, reliable, and uninterrupted access to your data at all times. It supports a wide range of applications, from...</p>
--	---	---	--	--

Bridging the Gap Survey

<p>AI-Boost by TechTonic</p> <p>AI-Boost is a cutting-edge artificial intelligence solution designed to optimize business operations. Powered by advanced machine learning algorithms, AI-Boost can analyze vast amounts of data to provide insightful predictions and recommendations. This enables businesses to make data-driven decisions, automate repetitive tasks, and improve overall efficiency. The solution can...</p>	<p>GeneCode by BioVital Dynamics</p> <p>GeneCode is a complete genomic analysis service that leverages advanced biotechnology to provide a comprehensive understanding of genetic material. From detailed DNA sequencing to in-depth gene-expression profiling, GeneCode delivers precise, actionable insights for researchers and healthcare providers. The service is able to decode the entire human genome in less than 24 hour...</p>	<p>Data-Driven Decision... by Data Drive Dynamics</p> <p>The Data-Driven Decision Suite is a comprehensive package of software solutions designed to empower businesses with actionable insights. The suite includes a range of tools for data collection, processing, analysis, and visualization. It leverages advanced machine learning algorithms to provide predictive analytics, enabling businesses to anticipate market trends and customer behavior. The suite also...</p>	<p>ProductTech AutoBots by ProductTech</p> <p>ProductTech AutoBots are advanced industrial robots designed for use in automobile manufacturing. They are equipped with state-of-the-art technology to ensure precision and efficiency in assembly line operations. With a wide range of motion and high-speed capabilities, these robots can perform complex tasks such as welding, painting, assembly, pick and place, and quality inspection. They are...</p>	<p>DataSphere by Data Drive Dynamics</p> <p>DataSphere is a comprehensive data management solution designed to simplify and streamline the process of collecting, processing, analyzing, and storing large volumes of data. It offers an intuitive user interface, enabling users to easily navigate through complex data landscapes. With its robust algorithms, DataSphere can quickly process and analyze data...</p>
--	---	--	--	---

Both search results feel equal

* 14. Results 3 of search: "Machine learning"

<p>ByteShield by ByteCortex</p> <p>ByteShield is a state-of-the-art cybersecurity solution developed by ByteCortex. It provides top-tier protection against a wide range of threats including malware, ransomware, phishing, and DDoS attacks. With its advanced machine learning algorithms, ByteShield can detect and neutralize threats before they cause harm. It comes with an easy-to-use interface, making it accessible to both IT...</p>	<p>CloudGuard Pro by TechNest</p> <p>CloudGuard Pro is an advanced and robust cybersecurity service from TechNest that safeguards your data in the cloud. It utilizes AI and machine learning to monitor and predict potential threats, offering proactive protection against cyber-attacks. CloudGuard Pro can be customized to cater to specific industry requirements and threats, making it an ideal solution f...</p>	<p>CyberShield by TechNova Dynamics</p> <p>CyberShield is a cutting-edge cybersecurity solution designed to protect businesses from a wide range of digital threats. This robust software suite includes antivirus, anti-malware, firewall, intrusion detection, and data encryption features. It uses advanced machine learning algorithms to detect and neutralize threats in real time, providing comprehensive protection for your digital assets...</p>	<p>CyberShield by TechNova Dynamics</p> <p>CyberShield is an advanced cybersecurity solution developed by TechNova Dynamics. This state-of-the-art software provides end-to-end protection for your business infrastructure from a wide range of cyber threats. CyberShield is equipped with machine learning algorithms that allow it to detect and neutralize threats in real time. Its user-friendly interface makes it easy to manage and monitor your organization's...</p>	<p>CloudNimbus by TechNimbus</p> <p>CloudNimbus is a robust and scalable cloud computing service offered by TechNimbus. This state-of-the-art service allows businesses to store, manage, and process data on a secure, remote server network. CloudNimbus is built on high-performance computing systems, ensuring fast, reliable, and uninterrupted access to your data at all times. It supports a wide range of applications, from...</p>
--	---	---	---	--

<p>DynaFlex Textile... by Fusion Dynamics</p> <p>The DynaFlex Textile Machine, a flagship product of Fusion Dynamics, is a versatile and high-performance textile manufacturing machine. Designed to handle diverse textile materials, this machine offers unparalleled efficiency and adaptability. It boasts advanced features such as automatic thread count adjustment and fabric tension control, ensuring optimal product quality at al...</p>	<p>OrbitSecure by TechOrbit</p> <p>OrbitSecure is a comprehensive cybersecurity solution designed to protect businesses from a wide range of cyber threats. It offers real-time threat detection, security incident management, and rapid response capabilities. With advanced machine learning algorithms, OrbitSecure can identify and neutralize threats before they can cause damage. This service is customizable according to...</p>	<p>CloudGuard Pro by TechNest</p> <p>CloudGuard Pro is an advanced and robust cybersecurity service from TechNest that safeguards your data in the cloud. It utilizes AI and machine learning to monitor and predict potential threats, offering proactive protection against cyber-attacks. CloudGuard Pro can be customized to cater to specific industry requirements and threats, making it an ideal solution f...</p>	<p>Cloud Optimizer by TechTonic</p> <p>Cloud Optimizer is a comprehensive cloud computing solution offered by TechTonic. It allows businesses to manage and optimize their cloud resources effectively. The solution includes features like automatic scaling, resource monitoring, cost management, and security compliance. The technical specifications of the product include compatibility with major cloud platforms...</p>	<p>AI-Boost by TechTonic</p> <p>AI-Boost is a cutting-edge artificial intelligence solution designed to optimize business operations. Powered by advanced machine learning algorithms, AI-Boost can analyze vast amounts of data to provide insightful predictions and recommendations. This enables businesses to make data-driven decisions, automate repetitive tasks, and improve overall efficiency. The solution can...</p>
---	---	---	--	--

Both search results feel equal

* 15. These search results are better than other search tools I have used.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

4 / 8

50%

Prev

Next

Powered by



See how easy it is to [create surveys and forms](#).

Bridging the Gap

Results Comparison of "Trading algorithms"

Trading algorithms are widely used in the financial sectors for their day-to-day trading. What if a business wanted to purchase such an algorithm?

* 16. Results 1 of search: "Trading algorithms"

<p>HexaSecure <small>by HexaSecure</small></p> <p>HexaSecure is an enterprise-grade cybersecurity solution designed to protect businesses from various cyber threats. It uses advanced AI algorithms to detect and neutralize threats in real-time. Key features include intrusion detection, data encryption, firewall configuration, and real-time threat alerts. HexaSecure can be customized to cater to specific business needs and is compatible with various...</p>	<p>Nebula SecureNet <small>by Nebula Networks</small></p> <p>Nebula SecureNet is a robust and secure networking solution designed to ensure seamless data transfer and communication across various platforms. By leveraging advanced encryption algorithms and security protocols, Nebula SecureNet offers businesses the peace of mind they need when handling sensitive data. The product can easily be customized to suit the specific needs of a business...</p>	<p>OrbitSecure <small>by TechSecure</small></p> <p>OrbitSecure is a comprehensive cybersecurity solution designed to protect businesses from a wide range of cyber threats. It offers real-time threat detection, security incident management, and rapid response capabilities. With advanced machine learning algorithms, OrbitSecure can identify and neutralize threats before they can cause damage. This service is customizable according to...</p>	<p>CyberGuard <small>by PulsePoint Solutions</small></p> <p>CyberGuard is a top-tier cybersecurity solution developed by PulsePoint Solutions. It is designed to provide robust and comprehensive security to businesses of all sizes. The software uses advanced AI algorithms to identify and neutralize threats before they can damage your system. CyberGuard also comes with a user-friendly interface that makes it easy for businesses to manage their security. Thi...</p>	<p>Cloud Optimizer <small>by TechTonic</small></p> <p>Cloud Optimizer is a comprehensive cloud computing solution offered by TechTonic. It allows businesses to manage and optimize their cloud resources effectively. The solution includes features like automatic scaling, resource monitoring, cost management, and security compliance. The technical specifications of the product include compatibility with major cloud platforms...</p>
<p>Matrix Money... <small>by Matrix Money</small></p> <p>A comprehensive wealth management service that provides clients with personalized financial planning and investment strategies. Our experienced advisors will guide you through the process of creating a diversified portfolio tailored to your financial goals and risk tolerance. This service includes regular portfolio reviews and adjustments to ensure optimal performance...</p>	<p>Matrix Money Manager <small>by Matrix Money</small></p> <p>Matrix Money Manager is a state-of-the-art digital banking solution that integrates seamlessly with your existing financial systems. It offers a range of features including real-time tracking of income and expenses, budget planning, and investment tracking. It also comes with an AI-powered financial advisor that offers personalized investment advice based on your financial goals and risk...</p>	<p>Matrix Invest Plus <small>by Matrix Money</small></p> <p>Matrix Invest Plus is a comprehensive, diversified investment portfolio solution that caters to various financial goals and risk appetites. It offers a wide array of investment options from equities, bonds, to alternative investments, strategically picked by our experienced portfolio managers. Depending on your risk tolerance, investment horizon, and financial objectives, you can choose...</p>	<p>Matrix Wealth... <small>by Matrix Money</small></p> <p>Matrix Wealth Management is a comprehensive financial planning service designed to help clients grow, manage, and protect their wealth. Our team of seasoned financial advisors will work closely with you to create a personalized investment strategy based on your financial goals, risk tolerance, and time horizon. We offer a wide range of investment options, including stocks, bonds, mutual funds, real...</p>	<p>Matrix Wealth... <small>by Matrix Money</small></p> <p>Our Matrix Wealth Management service is a comprehensive financial planning and investment solution designed to help our clients grow, manage, and protect their wealth. Our team of experienced wealth managers will work with you to understand your financial goals and create a personalized investment strategy. This service includes portfolio management, retirement...</p>

Both search results feel equal

* 17. Results 2 of search: "Trading algorithms"



Sorry, we couldn't find any results for your search.

Bridging the Gap Survey

HexaSecure <small>by HexaSecure</small> HexaSecure is an enterprise-grade cybersecurity solution designed to protect businesses from various cyber threats. It uses advanced AI algorithms to detect and neutralize threats in real-time. Key features include intrusion detection, data encryption, firewall configuration, and real-time threat alerts. HexaSecure can be customized to cater to specific business needs and is compatible with various...	Nebula SecureNet <small>by Nebula Networks</small> Nebula SecureNet is a robust and secure networking solution designed to ensure seamless data transfer and communication across various platforms. By leveraging advanced encryption algorithms and security protocols, Nebula SecureNet offers businesses the peace of mind they need when handling sensitive data. The product can easily be customized to suit the specific needs of a business...	OrbitSecure <small>by TechSecure</small> OrbitSecure is a comprehensive cybersecurity solution designed to protect businesses from a wide range of cyber threats. It offers real-time threat detection, security incident management, and rapid response capabilities. With advanced machine learning algorithms, OrbitSecure can identify and neutralize threats before they can cause damage. This service is customizable according to...	CyberGuard <small>by PulseGuard Solutions</small> CyberGuard is a top-tier cybersecurity solution developed by PulseGuard Solutions. It is designed to provide robust and comprehensive security to businesses of all sizes. The software uses advanced AI algorithms to identify and neutralize threats before they can damage your system. CyberGuard also comes with a user-friendly interface that makes it easy for businesses to manage their security. TH...	Cloud Optimizer <small>by TechTonic</small> Cloud Optimizer is a comprehensive cloud computing solution offered by TechTonic. It allows businesses to manage and optimize their cloud resources effectively. The solution includes features like automatic scaling, resource monitoring, cost management, and security compliance. The technical specifications of the product include compatibility with major cloud platforms...
---	--	---	--	---

Both search results feel equal

* 18. Results 3 of search: "Trading algorithms"

Matrix Money... <small>by MoneyMatrix</small> A comprehensive wealth management service that provides clients with personalized financial planning and investment strategies. Our experienced advisors will guide you through the process of creating a diversified portfolio tailored to your financial goals and risk tolerance. This service includes regular portfolio reviews and adjustments to ensure optimal performance...	Matrix Money Manager <small>by MoneyMatrix</small> Matrix Money Manager is a state-of-the-art digital banking solution that integrates seamlessly with your existing financial systems. It offers a range of features including real-time tracking of income and expenses, budget planning, and investment tracking. It also comes with an AI-powered financial advisor that offers personalized investment advice based on your financial goals and risk...	Matrix Invest Plus <small>by MoneyMatrix</small> Matrix Invest Plus is a comprehensive, diversified investment portfolio solution that caters to various financial goals and risk appetites. It offers a wide array of investment options from equities, bonds, to alternative investments, strategically picked by our experienced portfolio managers. Depending on your risk tolerance, investment horizon, and financial objectives, you can choose...	Matrix Wealth... <small>by MoneyMatrix</small> Matrix Wealth Management is a comprehensive financial planning service designed to help clients grow, manage, and protect their wealth. Our team of seasoned financial advisors will work closely with you to create a personalized investment strategy based on your financial goals, risk tolerance, and time horizon. We offer a wide range of investment options, including stocks, bonds, mutual funds, real...	Matrix Wealth... <small>by MoneyMatrix</small> Our Matrix Wealth Management service is a comprehensive financial planning and investment solution designed to help our clients grow, manage, and protect their wealth. Our team of experienced wealth managers will work with you to understand your financial goals and create a personalized investment strategy. This service includes portfolio management, retirement...
--	---	--	--	--



Sorry, we couldn't find any results for your search.

Both search results feel equal

* 19. These search results are better than other search tools I have used.

Strongly agree

Agree

Neither agree nor disagree

Disagree

Strongly disagree

5 / 8

62%

Prev

Next

Powered by



See how easy it is to [create surveys and forms](#).

Bridging the Gap

Results Comparison of "Sustainable energy storage"

As a community we are slowly moving to more sustainable energy production. What if I have just installed a set of solar panels on my new factory and was now looking for solutions to store this energy?

* 20. Results 1 of search: "Sustainable energy storage"

<p>Voltus SolarPro <small>by Voltus Solar</small></p> <p>Voltus SolarPro is a top-tier solar energy solution, designed for both residential and commercial use. It harnesses the power of the sun to provide clean, renewable energy, thereby reducing reliance on fossil fuels and contributing to a more sustainable future. With an efficiency rate of over 22%, our solar panels are among the most efficient on the market. Technical Specifications: - Power...</p>	<p>SolarFlux Power... <small>by SolarFlux Power</small></p> <p>SolarFlux is an advanced solar power generation system designed to optimize energy use and minimize environmental impact. It is equipped with high-efficiency photovoltaic cells that convert solar energy into electricity with an efficiency rate of over 20%. The system comes with a smart energy management system that allows for real-time monitoring and control of power generation and...</p>	<p>Quantum Solar Power... <small>by Quantum Solar</small></p> <p>Our Quantum Solar Power Systems are state-of-the-art renewable energy solutions designed to optimize energy use and minimize environmental impact. These systems utilize advanced photovoltaic technology to convert sunlight into electricity, providing a reliable and sustainable source of power. The systems are scalable, allowing for installation in residential, commercial, and industrial...</p>	<p>Solaris Residential... <small>by Solaris Energy</small></p> <p>Our Solaris Residential Panels are a high-performance solution designed for residential use. These panels are made from high-quality, durable materials that can withstand harsh weather conditions. They are easy to install and require minimal maintenance. The panels come in various sizes and capacities, ranging from 100W to 400W, to cater to different energy needs. They are equipped with...</p>	<p>SolarFlux HomeStar <small>by SolarFlux</small></p> <p>The SolarFlux HomeStar is a residential solar power system designed for maximum efficiency and reliability. This system utilizes our high-performance solar panels, which are known for their superior efficiency, durability, and adaptability to diverse environmental conditions. The HomeStar system comes in various capacities, ranging from 5 kW to 10 kW, allowing customers to choose the system that...</p>
--	---	---	---	---

<p>Sustainable Investmen... <small>by Monetary Maven</small></p> <p>This is an investment portfolio service that is conscious of environmental preservation. Our team of experts will curate a diverse range of investment options that are both profitable and eco-friendly. These options could range from renewable energy companies to organic farming businesses. This will not only ensure a good return on investment but also contribute to the well-being of our planet. Pricing is...</p>	<p>Sustainable Investmen... <small>by Fiscal Harmony</small></p> <p>Our Sustainable Investment Portfolio is a curated investment option that allows clients to invest in environmentally friendly and socially responsible businesses. This portfolio is meticulously crafted by our expert advisors, focusing on companies that are leading the way in sustainable practices. These investments not only offer competitive returns but also contribute to a greener future. The...</p>	<p>EcoTech Smart Fridge <small>by Prodigal Production</small></p> <p>Our EcoTech Smart Fridge is a revolution in food preservation technology. It's equipped with state-of-the-art sensors to maintain optimal temperature and humidity levels, ensuring your food stays fresh for longer. The fridge also features an intelligent inventory system that keeps track of your food items and alerts you when they're about to expire. Additionally, it can suggest recipes based on the...</p>	<p>EcoTech Smart Fridge <small>by Prodigal Production</small></p> <p>The EcoTech Smart Fridge is an innovative product from Prodigal Production that redefines the concept of refrigeration. This high-tech refrigerator is equipped with advanced features like energy-efficient cooling, smart temperature control, and an interactive touch screen panel. It also has a built-in Wi-Fi connection and is compatible with smart home systems. The fridge has a sleek and modern design...</p>	<p>HydroPro Energy... <small>by AquaBank</small></p> <p>HydroPro Energy Systems are our flagship hydroelectric power plants designed to harness the power of water to generate clean, renewable energy. These systems are built with cutting-edge technology ensuring high efficiency and minimal environmental impact. Our HydroPro systems are available in different capacities, ranging from small-scale plants suitable for rural areas to large-scal...</p>
--	--	--	--	--

Both search results feel equal

* 21. Results 2 of search: "Sustainable energy storage"

<p>Sustainable Investmen... by Monitory Maren</p> <p>This is an investment portfolio service that is conscious of environmental preservation. Our team of experts will curate a diverse range of investment options that are both profitable and eco-friendly. These options could range from renewable energy companies to organic farming businesses. This will not only ensure a good return on investment but also contribute to the well-being of our planet. Pricing is...</p>	<p>Sustainable Investmen... by Fiscal Harmony</p> <p>Our Sustainable Investment Portfolio is a curated investment option that allows clients to invest in environmentally friendly and socially responsible businesses. This portfolio is meticulously crafted by our expert advisors, focusing on companies that are leading the way in sustainable practices. These investments not only offer competitive returns but also contribute to a greener future. The...</p>	<p>EcoTech Smart Fridge by Prodigal Production</p> <p>Our EcoTech Smart Fridge is a revolution in food preservation technology. It's equipped with state-of-the-art sensors to maintain optimal temperature and humidity levels, ensuring your food stays fresh for longer. The fridge also features an intelligent inventory system that keeps track of your food items and alerts you when they're about to expire. Additionally, it can suggest recipes based on the...</p>	<p>EcoTech Smart Fridge by Prodigal Production</p> <p>The EcoTech Smart Fridge is an innovative product from Prodigal Production that redefines the concept of refrigeration. This high-tech refrigerator is equipped with advanced features like energy-efficient cooling, smart temperature control, and an interactive touch screen panel. It also has a built-in Wi-Fi connection and is compatible with smart home systems. The fridge has a sleek and modern design...</p>	<p>HydroPro Energy-- by AquaBank</p> <p>HydroPro Energy Systems are our flagship hydroelectric power plants designed to harness the power of water to generate clean, renewable energy. These systems are built with cutting-edge technology ensuring high efficiency and minimal environmental impact. Our HydroPro systems are available in different capacities, ranging from small-scale plants suitable for rural areas to large-scal...</p>
---	---	---	---	--

<p>CloudFusion by TechFusion Dynamics</p> <p>CloudFusion is our premium cloud computing service designed to offer businesses the flexibility and scalability they need. With CloudFusion, you can easily store, manage, and process data on our secure, global network of servers. The service includes features like auto-scaling, data analytics, and machine learning capabilities. You also have the option to choose from public, private, or hybrid cloud models based on your...</p>	<p>Solaris Residential... by Solaris Solutions</p> <p>Our Solaris Residential Panels are a high-performance solution designed for residential use. These panels are made from high-quality, durable materials that can withstand harsh weather conditions. They are easy to install and require minimal maintenance. The panels come in various sizes and capacities, ranging from 100W to 400W, to cater to different energy needs. They are equipped with...</p>	<p>SolarFlux HomeStar by SolarFlux</p> <p>The SolarFlux HomeStar is a residential solar power system designed for maximum efficiency and reliability. This system utilizes our high-performance solar panels, which are known for their superior efficiency, durability, and adaptability to diverse environmental conditions. The HomeStar system comes in various capacities, ranging from 5 kW to 10 kW, allowing customers to choose the system that...</p>	<p>SolarFlux Power... by Quantum Cascade</p> <p>SolarFlux is an advanced solar power generation system designed to optimize energy use and minimize environmental impact. It is equipped with high-efficiency photovoltaic cells that convert solar energy into electricity with an efficiency rate of over 20%. The system comes with a smart energy management system that allows for real-time monitoring and control of power generation and...</p>	<p>Quantum Solar Power... by Quantum Cascade</p> <p>Our Quantum Solar Power Systems are state-of-the-art renewable energy solutions designed to optimize energy use and minimize environmental impact. These systems utilize advanced photovoltaic technology to convert sunlight into electricity, providing a reliable and sustainable source of power. The systems are scalable, allowing for installation in residential, commercial, and industrial...</p>
--	---	--	--	--

Both search results feel equal

* 22. Results 3 of search: "Sustainable energy storage"

<p>CloudFusion by TechFusion Dynamics</p> <p>CloudFusion is our premium cloud computing service designed to offer businesses the flexibility and scalability they need. With CloudFusion, you can easily store, manage, and process data on our secure, global network of servers. The service includes features like auto-scaling, data analytics, and machine learning capabilities. You also have the option to choose from public, private, or hybrid cloud models based on your...</p>	<p>Solaris Residential... by Solaris Solutions</p> <p>Our Solaris Residential Panels are a high-performance solution designed for residential use. These panels are made from high-quality, durable materials that can withstand harsh weather conditions. They are easy to install and require minimal maintenance. The panels come in various sizes and capacities, ranging from 100W to 400W, to cater to different energy needs. They are equipped with...</p>	<p>SolarFlux HomeStar by SolarFlux</p> <p>The SolarFlux HomeStar is a residential solar power system designed for maximum efficiency and reliability. This system utilizes our high-performance solar panels, which are known for their superior efficiency, durability, and adaptability to diverse environmental conditions. The HomeStar system comes in various capacities, ranging from 5 kW to 10 kW, allowing customers to choose the system that...</p>	<p>SolarFlux Power... by Quantum Cascade</p> <p>SolarFlux is an advanced solar power generation system designed to optimize energy use and minimize environmental impact. It is equipped with high-efficiency photovoltaic cells that convert solar energy into electricity with an efficiency rate of over 20%. The system comes with a smart energy management system that allows for real-time monitoring and control of power generation and...</p>	<p>Quantum Solar Power... by Quantum Cascade</p> <p>Our Quantum Solar Power Systems are state-of-the-art renewable energy solutions designed to optimize energy use and minimize environmental impact. These systems utilize advanced photovoltaic technology to convert sunlight into electricity, providing a reliable and sustainable source of power. The systems are scalable, allowing for installation in residential, commercial, and industrial...</p>
--	---	--	--	--

<p>Voltus SolarPro by Voltus Energy</p> <p>Voltus SolarPro is a top-tier solar energy solution, designed for both residential and commercial use. It harnesses the power of the sun to provide clean, renewable energy, thereby reducing reliance on fossil fuels and contributing to a more sustainable future. With an efficiency rate of over 22%, our solar panels are among the most efficient on the market. Technical Specifications: - Power...</p>	<p>SolarFlux Power... by Quantum Cascade</p> <p>SolarFlux is an advanced solar power generation system designed to optimize energy use and minimize environmental impact. It is equipped with high-efficiency photovoltaic cells that convert solar energy into electricity with an efficiency rate of over 20%. The system comes with a smart energy management system that allows for real-time monitoring and control of power generation and...</p>	<p>Quantum Solar Power... by Quantum Cascade</p> <p>Our Quantum Solar Power Systems are state-of-the-art renewable energy solutions designed to optimize energy use and minimize environmental impact. These systems utilize advanced photovoltaic technology to convert sunlight into electricity, providing a reliable and sustainable source of power. The systems are scalable, allowing for installation in residential, commercial, and industrial...</p>	<p>Solaris Residential... by Solaris Solutions</p> <p>Our Solaris Residential Panels are a high-performance solution designed for residential use. These panels are made from high-quality, durable materials that can withstand harsh weather conditions. They are easy to install and require minimal maintenance. The panels come in various sizes and capacities, ranging from 100W to 400W, to cater to different energy needs. They are equipped with...</p>	<p>SolarFlux HomeStar by SolarFlux</p> <p>The SolarFlux HomeStar is a residential solar power system designed for maximum efficiency and reliability. This system utilizes our high-performance solar panels, which are known for their superior efficiency, durability, and adaptability to diverse environmental conditions. The HomeStar system comes in various capacities, ranging from 5 kW to 10 kW, allowing customers to choose the system that...</p>
--	--	--	---	--

Both search results feel equal

* 23. These search results are better than other search tools I have used.

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

6 / 8

75%

Prev

Next

Powered by



See how easy it is to [create surveys and forms](#).

Bridging the Gap

Results Comparison of "A gearbox for speedy cars under 1000\$"

I'm a manufacturer of race cars and need a new gearbox on a tight budget.

* 24. Results 1 of search: "A gearbox for speedy cars under 1000\$"



Sorry, we couldn't find any results for your search.

<input type="radio"/>	<p>Customized Gearbox <small>by TopGear</small></p> <p>The Customized Gearbox is a high-quality, durable component designed for use in various automotive applications. This product is crafted with precision and attention to detail, ensuring optimal performance and longevity. The gearbox can be customized according to the client's specifications, including size, gear ratios, and material. The average price range for this product varies depending on the level.</p>	<p>Quantum Gearbox <small>by Quantum</small></p> <p>A revolutionary product from our automotive line, the Quantum Gearbox boasts cutting-edge technology and unbeatable performance. This highly accurate and efficient gear system guarantees smoother gear shifts and a superior driving experience. It is compatible with a wide range of vehicles and can be customized to suit specific requirements. The Quantum Gearbox is manufactured with eco-friendly materials.</p>	<p>Effusion - High Precision... <small>by Quantum</small></p> <p>Effusion is our premium line of high-precision gearboxes designed for automobile and heavy machinery applications. Manufactured with state-of-the-art technology, these gearboxes offer superior performance and durability. They feature high-quality materials for increased lifespan and robustness. The technical specifications include a torque range of 100-500 Nm, rotational speed up to 1000...</p>	<p>Quantum Auto Gears <small>by Quantum</small></p> <p>Quantum Auto Gears are high-quality automotive components designed with precision and durability in mind. These gears are manufactured using advanced CNC machines, ensuring exact specifications and consistent performance. Our auto gears are suitable for a wide range of vehicles, from passenger cars to heavy-duty trucks. They come in various sizes, ratios, and types, catering to...</p>	<p>Prodigy Precision... <small>by Quantum</small></p> <p>Our Prodigy Precision Automotive Parts are engineered with utmost precision to ensure high performance and durability. These parts are designed to withstand the rigors of high-speed and heavy-duty vehicles, providing you with a reliable solution that you can trust. Our products are compatible with a wide range of vehicle models, ensuring a perfect fit for your specific needs. We offer a variety of...</p>
-----------------------	---	--	---	---	---



Both search results feel equal

* 25. Results 2 of search: "A gearbox for speedy cars under 1000\$"



<p>QuantumDrive... <small>by Quantum</small></p> <p>The QuantumDrive Automotive Transmission is a state-of-the-art transmission system designed to deliver unrivalled performance and efficiency. It features a seamless 8-speed automatic gearbox that ensures smooth gear transitions, providing drivers with a comfortable and enjoyable driving experience. The QuantumDrive is also equipped with our exclusive fuel-saving technology, whic...</p>	<p>Customized Gearbox <small>by TopGear</small></p> <p>The Customized Gearbox is a high-quality, durable component designed for use in various automotive applications. This product is crafted with precision and attention to detail, ensuring optimal performance and longevity. The gearbox can be customized according to the client's specifications, including size, gear ratios, and material. The average price range for this product varies depending on the level.</p>	<p>Quantum Gearbox <small>by Quantum</small></p> <p>A revolutionary product from our automotive line, the Quantum Gearbox boasts cutting-edge technology and unbeatable performance. This highly accurate and efficient gear system guarantees smoother gear shifts and a superior driving experience. It is compatible with a wide range of vehicles and can be customized to suit specific requirements. The Quantum Gearbox is manufactured with eco-friendly materials.</p>	<p>VersaGear <small>by Quantum</small></p> <p>The VersaGear is a state-of-the-art gearbox part designed for high-performance vehicles. Manufactured with precision and accuracy, this product guarantees smooth vehicle operation and increased longevity. With its modern design and durable material, it offers superior performance and durability that sets it apart from conventional gearbox parts. Technical Specifications: - Material...</p>	<p>CircuitSphere S-1000 <small>by Quantum</small></p> <p>The CircuitSphere S-1000 is a high-performance, multi-layered circuit board designed for high-powered electronics. It is built with state-of-the-art machinery and undergoes strict quality control to ensure it meets the highest industry standards. The S-1000 comes with customizable options, where customers have the flexibility to choose from various sizes and technical specifications to match their specific need...</p>
---	---	--	--	---



Both search results feel equal

* 26. Results 3 of search: "A gearbox for speedy cars under 1000\$"



<p>QuantumDrive... <small>by Prodigy Motors</small></p> <p>The QuantumDrive Automotive Transmission is a state-of-the-art transmission system designed to deliver unrivaled performance and efficiency. It features a seamless 8-speed automatic gearbox that ensures smooth gear transitions, providing drivers with a comfortable and enjoyable driving experience. The QuantumDrive is also equipped with our exclusive fuel-saving technology, whic...</p>	<p>Customized Gearbox <small>by Forge & Fabricate</small></p> <p>The Customized Gearbox is a high-quality, durable component designed for use in various automotive applications. This product is crafted with precision and attention to detail, ensuring optimal performance and longevity. The gearbox can be customized according to the client's specifications, including size, gear ratios, and material. The average price range for this product varies depending on the leve...</p>	<p>Quantum Gearbox <small>by Quantum Forge</small></p> <p>A revolutionary product from our automotive line, the Quantum Gearbox boasts cutting-edge technology and unbeatable performance. This highly accurate and efficient gear system guarantees smoother gear shifts and a superior driving experience. It is compatible with a wide range of vehicles and can be customized to suit specific requirements. The Quantum Gearbox is manufactured with eco-friendly materials...</p>	<p>VersaGear <small>by FusionFlow</small></p> <p>The VersaGear is a state-of-the-art gearbox part designed for high-performance vehicles. Manufactured with precision and accuracy, this product guarantees smooth vehicle operation and increased longevity. With its modern design and durable material, it offers superior performance and durability that sets it apart from conventional gearbox parts. Technical Specifications: - Material...</p>	<p>CircuitSphere S-1000 <small>by CircuitSphere</small></p> <p>The CircuitSphere S-1000 is a high-performance, multi-layered circuit board designed for high-powered electronics. It is built with state-of-the-art machinery and undergoes strict quality control to ensure it meets the highest industry standards. The S-1000 comes with customizable options, where customers have the flexibility to choose from various sizes and technical specifications to match their specific need...</p>
---	--	--	---	---



<p>Customized Gearbox <small>by Forge & Fabricate</small></p> <p>The Customized Gearbox is a high-quality, durable component designed for use in various automotive applications. This product is crafted with precision and attention to detail, ensuring optimal performance and longevity. The gearbox can be customized according to the client's specifications, including size, gear ratios, and material. The average price range for this product varies depending on the leve...</p>	<p>Quantum Gearbox <small>by Quantum Forge</small></p> <p>A revolutionary product from our automotive line, the Quantum Gearbox boasts cutting-edge technology and unbeatable performance. This highly accurate and efficient gear system guarantees smoother gear shifts and a superior driving experience. It is compatible with a wide range of vehicles and can be customized to suit specific requirements. The Quantum Gearbox is manufactured with eco-friendly materials...</p>	<p>Effusion - High Precisi... <small>by EffusionFlow</small></p> <p>Effusion is our premium line of high-precision gearboxes designed for automobile and heavy machinery applications. Manufactured with state-of-the-art technology, these gearboxes offer superior performance and durability. They feature high-quality materials for increased lifespan and robustness. The technical specifications include a torque range of 100-500 Nm, rotational speed up to 10,000...</p>	<p>Quantum Auto Gears <small>by Quantum Forge</small></p> <p>Quantum Auto Gears are high-quality automotive components designed with precision and durability in mind. These gears are manufactured using advanced CNC machines, ensuring exact specifications and consistent performance. Our auto gears are suitable for a wide range of vehicles, from passenger cars to heavy-duty trucks. They come in various sizes, ratios, and types, catering to...</p>	<p>Prodigy Precision... <small>by Prodigy Motors</small></p> <p>Our Prodigy Precision Automotive Parts are engineered with utmost precision to ensure high performance and durability. These parts are designed to withstand the rigors of high-speed and heavy-duty vehicles, providing you with a reliable solution that you can trust. Our products are compatible with a wide range of vehicle models, ensuring a perfect fit for your specific needs. We offer a variety of...</p>
--	--	--	---	--

Both search results feel equal

* 27. These search results are better than other search tools I have used.

Strongly agree

Agree

Neither agree nor disagree

Disagree

Strongly disagree

7 / 8

88%

Prev

Next

Powered by



See how easy it is to [create surveys and forms](#).

Bridging the Gap

Future Collaboration

* 28. Would you be open to participating in a follow-up interview with regards to creating a platform to connecting businesses?

Yes

No

29. Email

Email address

8 / 8

100%

Prev

Done

Powered by



See how easy it is to [create surveys and forms](#).