

Semantic Search in AI for Automated Matching of Customer Purchase Orders with Supplier Sales Orders

Name: Mahmoud Mohammed Alra'e

Abstract

This explores the application of semantic search within artificial intelligence (AI) to solve the problem of matching item descriptions in customer purchase orders to corresponding items in a database. Semantic search interprets the intent and context of text, rather than merely matching keywords, and thus promises to improve the accuracy and efficiency of product matching. Retrieval augmented generation, knowledge graph integration, multilingual retrieval and the wider technological context of AI-driven search systems. The review provides concise summaries, critical analysis, connections, conflicts and a synthesis matrix summarizing the key aspects of each study.

In Fact, Semantic search represents a significant advancement in retrieving information by focusing on contextual meaning rather than exact keyword matches. In procurement environments, users often provide vague, incomplete or inconsistent item descriptions, making traditional search methods unreliable. This proposal examines how transformer based semantic models such as SBERT, MPNet, and DPR can enhance automated matching of purchase order descriptions.

The research aims to develop a structured framework that improves retrieval accuracy, reduces manual correction, and supports scalable procurement workflows.

Introduction

As organizations continue to digitize their procurement processes, ensuring that item descriptions are accurately matched to the correct products has become more important than ever. In reality, customers rarely use consistent or complete wording when submitting purchase orders. They may rely on abbreviations, informal descriptions, or unfamiliar terminology, which makes it difficult for traditional keyword based search systems to interpret their requests correctly. Because these systems depend heavily on exact word matching, even small variations in phrasing can lead to incorrect or missed results.

Semantic search offers a promising alternative by focusing on the meaning behind the words rather than the words themselves. By understanding the user's intent, semantic models can handle inconsistent, vague, or non standard descriptions more effectively, ultimately improving the accuracy and reliability of item matching. This introduction sets the stage for exploring why semantic search is becoming increasingly valuable in procurement environments and how it can support the development of more intelligent and efficient retrieval frameworks.

Background

Semantic search has come a long way, evolving from early systems that relied on predefined rules and ontologies to today's advanced models that use neural networks to understand the context behind words. The introduction of transformer architectures, like BERT and its variations, has revolutionized how we measure semantic similarity, allowing for much more accurate and nuanced text interpretation.

In procurement, traditional search methods often struggle due to the variability in terminology, spelling mistakes, and incomplete or unclear item descriptions. These challenges make it difficult for basic search systems to find the right matches. Unlike traditional methods, semantic search focuses on understanding the underlying meaning of the words, rather than just matching exact phrases. By doing so, it addresses the common pitfalls of inconsistent descriptions and offers a much more reliable way to search for items.

This section provides an overview of how semantic search works, why it's so relevant for procurement, and the specific challenges that this research aims to solve.

Problem Statement

As more companies shift their procurement and inventory operations into digital platforms, a recurring challenge continues to emerge: customers rarely describe products in a consistent or predictable way. One customer may provide the full technical name of an item, another might include only part of it or mention a brand, while others use vague descriptions like “blue cable” or “the usual filter.” Spelling mistakes, missing details, and informal language make these descriptions even harder for automated systems to understand.

Traditional keyword based search tools struggle with this level of variation because they rely on exact or near exact word matches. When the system cannot confidently identify the intended item, human staff must step in to correct or interpret the request. This slows down the workflow, increases the likelihood of errors, and becomes even more problematic as the number of items in a catalog grows.

To address these issues, many organizations have started exploring AI driven semantic search. Rather than matching words directly, semantic search aims to understand what the user *means*. However, even these advanced methods are not without limitations. Many semantic models still have difficulty processing incomplete or messy descriptions, adapting to evolving terminology, or handling multilingual input. Some approaches depend on large, domain specific datasets, which most organizations do not have. Models that perform well in laboratory settings may also fail to meet the demands of fast paced, real world procurement environments.

This disconnect between the capabilities of current tools and the needs of actual business workflows has real consequences: delays, increased costs, and a continued reliance on manual corrections. Because accurate item matching is essential to automated procurement, there is a clear need for a more adaptable approach one that can handle the many different ways people describe items. To develop such a solution, it is important to first understand the strengths and weaknesses of existing methods and identify what can realistically be improved within business environments

Research Objectives

1. Examine the weaknesses in current semantic search methods and identify the main reasons they fail when dealing with real purchase order descriptions.
2. Review and compare linguistic, ontology based, embedding based and hybrid approaches to understand how each method contributes to semantic similarity.
3. Develop a conceptual framework that brings together the most practical elements from these methods and adapts them to item matching tasks.
4. Test the proposed framework using a variety of item descriptions and measure its performance using precision, recall, ranking scores, and other standard metrics.
5. Recommend practical, realistic strategies for organizations that want to use semantic search to reduce manual corrections and improve matching accuracy

Research Scope

This research is centered on improving semantic retrieval for item descriptions in procurement settings. It focuses on preparing and cleaning textual data, evaluating transformer based embedding models, and developing a semantic matching framework designed specifically for procurement workflows. The study aims to assess how accurately these models retrieve items, how well they rank relevant results, and how adaptable they are when faced with inconsistent or varied item descriptions.

However, the scope of this research is intentionally limited. It does not involve building or deploying a real time procurement system, connecting to proprietary company databases, or creating production ready user interfaces. The study also does not explore multimodal retrieval methods such as image or voice search, perform extensive multilingual analysis beyond the available datasets, or address full scale enterprise deployment. These areas fall outside the boundaries of this work and may be considered for future research

Literature Review

Traditional keyword search systems work reasonably well when users enter clean and consistent text. Unfortunately, purchase order descriptions rarely look like that. Customers may use older terminology, abbreviations, or their own phrasing, and the gap between their descriptions and the official product names creates difficulties for simple keyword matching. This problem becomes especially visible when the text includes spelling errors or informal wording.

Researchers have explored different ways to address this. One early idea was query expansion, where the system adds related terms to the original query in hopes of finding a match. Although this sometimes improves the results, it can also introduce items that have little to do with the user's actual request. Another well known approach is the use of ontologies structured collections of concepts and their relationships. Ontologies can be very helpful in fields where terminology is stable, but they require significant effort to build and maintain. They also struggle in environments where products change quickly or where multiple languages are involved.

More recent work has focused on neural embedding models, which capture the meaning of text by representing words or sentences as vectors. Models like Sentence-BERT, MPNet, and ColBERT have shown strong performance in ranking and similarity tasks, especially when item descriptions vary in style or structure. However, embedding typically require substantial computational resources, and they don't always perform well if trained on data that doesn't match the organization's domain closely.

Some studies have explored hybrid approaches, such as retrieval augmented generation (RAG). These systems pull in potentially relevant information and then use a generative model to interpret it. While promising in question answering and summarization tasks, these methods can introduce errors or add unsupported details problems that make them risky for tasks like item matching, where accuracy is essential.

Despite their differences, the research points to several common themes. Techniques that incorporate context whether through structured domain knowledge, embedding, or hybrid strategies tend to outperform simple keyword matching. Many studies also suggest that combining multiple approaches can produce more reliable results. With the increasing need to handle multilingual customer input, researchers have also highlighted the importance of models trained on diverse, multilingual datasets, which usually perform more consistently across regions and suppliers.

Research Methodology

This study follows a structured, multi phase methodology designed to investigate and evaluate semantic search models for automating the matching of purchase order descriptions. The approach is modeled after established research practices in information retrieval and aims to develop a reliable, domain specific semantic matching framework supported by empirical evidence.

Phase 1: Literature Review

The first phase involves a detailed and systematic review of scholarly literature related to semantic search, neural embedding, contextual similarity, and procurement automation. Academic databases such as IEEE Xplore, ACM Digital Library, SpringerLink and Google Scholar serve as principal sources. The goal of this phase is to map the evolution of semantic search models, identify their strengths and limitations and understand how current research trends may be adapted to procurement specific contexts. Comparative synthesis will be used to categorize existing models based on their suitability for domain adaptation, retrieval efficiency, and semantic robustness. By the end of this phase a theoretical foundation is established to guide subsequent experimentation.

Phase 2: Data Collection and Model Preparation

This phase focuses on gathering procurement related item descriptions and catalog data from publicly available datasets or simulated procurement corpora. Data preprocessing includes text normalization, removal of noise, tokenization, spelling correction and generation of synthetic variations to mimic real world purchase order variability. Transformer based models such as SBERT, MPNet, DPR and ColBERT are selected for evaluation.

Each model undergoes domain adaptation through fine tuning or embedding calibration to ensure sensitivity to procurement terminology. The objective of this phase is to create a standardized experimental pipeline capable of comparing models fairly across diverse retrieval conditions.

Phase 3: Quantitative Evaluation

The third phase employs quantitative metrics to evaluate model performance. Metrics include precision@k, recall@k, Mean Reciprocal Rank (MRR), cosine similarity, and nDCG. These metrics capture both ranking quality and semantic similarity. Evaluation is conducted on test sets containing ambiguous, incomplete and informal item descriptions. Error analysis is also performed to identify patterns in mismatched predictions. Findings from this phase provide critical insights into the strengths and weaknesses of each model and highlight areas where domain specific adaptation may be necessary.

Phase 4: Framework Development and Validation

The final phase integrates findings from previous phases into a practical semantic matching framework. This framework includes guidelines for preprocessing item descriptions, selecting embedding models, defining similarity thresholds, and integrating semantic search into procurement workflows. Validation occurs through scenario based testing and expert review, evaluating the framework's applicability in real procurement settings. The goal is to produce a scalable, efficient, and accurate retrieval system that reduces manual intervention and improves decision making consistency.

Ethical Considerations

Ethical considerations are an important part of research involving semantic search models, especially in procurement environments where item descriptions may contain proprietary or sensitive information. To protect confidentiality, all datasets used in this study will be carefully anonymized. This means that company names, product identifiers, and any form of user data will be removed to ensure that no sensitive information is revealed at any stage of the research.

Another ethical concern relates to the potential for semantic models to unintentionally reproduce or amplify biases present in the data they were trained on. To address this risk, model outputs will be monitored and evaluated throughout the study. Any patterns of bias or unfair retrieval behavior will be documented, and strategies for reducing or mitigating these issues will be explored.

Transparency is also a key ethical requirement. All decisions made during the research process—including the selection of models, fine tuning approaches, and evaluation methods—will be clearly reported. The study will avoid overstating the capabilities of semantic search models and will acknowledge their limitations to ensure that findings are interpreted responsibly and accurately.

Limitations

This research may be affected by several limitations. First, procurement data differs significantly across industries, which means that a model fine tuned for one domain may not perform as well in another. Although domain adaptation techniques will be applied, it is unlikely that any single model will generalize perfectly across all sectors.

Second, the study may face challenges related to data quality and availability. Many organizations have item descriptions that are incomplete, inconsistent, abbreviated, or filled with noise. Such variability can make it difficult for even advanced semantic models to learn effectively, especially when large, high quality datasets are not readily accessible.

Third, computational constraints may limit the scale of experiments. Transformer based architectures require substantial memory and processing power, which may influence decisions about how extensively models can be fine tuned or how large the embedding dimensions can be.

Finally, while human evaluation provides valuable insights, it also introduces an element of subjectivity. Different evaluators may interpret ambiguous item descriptions differently, which can affect the consistency and reliability of qualitative assessments of retrieval performance.

Expected Outcomes

This research is expected to generate several meaningful outcomes. First, it aims to show clear improvements in retrieval accuracy by using transformer based semantic models, especially when dealing with ambiguous, incomplete or inconsistently written purchase order descriptions.

Second, the study will deliver a semantic matching framework designed specifically for procurement systems. This framework will offer practical guidance on text preprocessing, selecting appropriate embedding models, setting similarity thresholds, and evaluating retrieval performance making it a useful resource for organizations working to enhance their search and matching processes.

Third, the research will contribute to academic understanding by highlighting the unique challenges of semantic retrieval in enterprise environments. It will provide comparative insights into how different embedding based models perform when applied to domain specific procurement data.

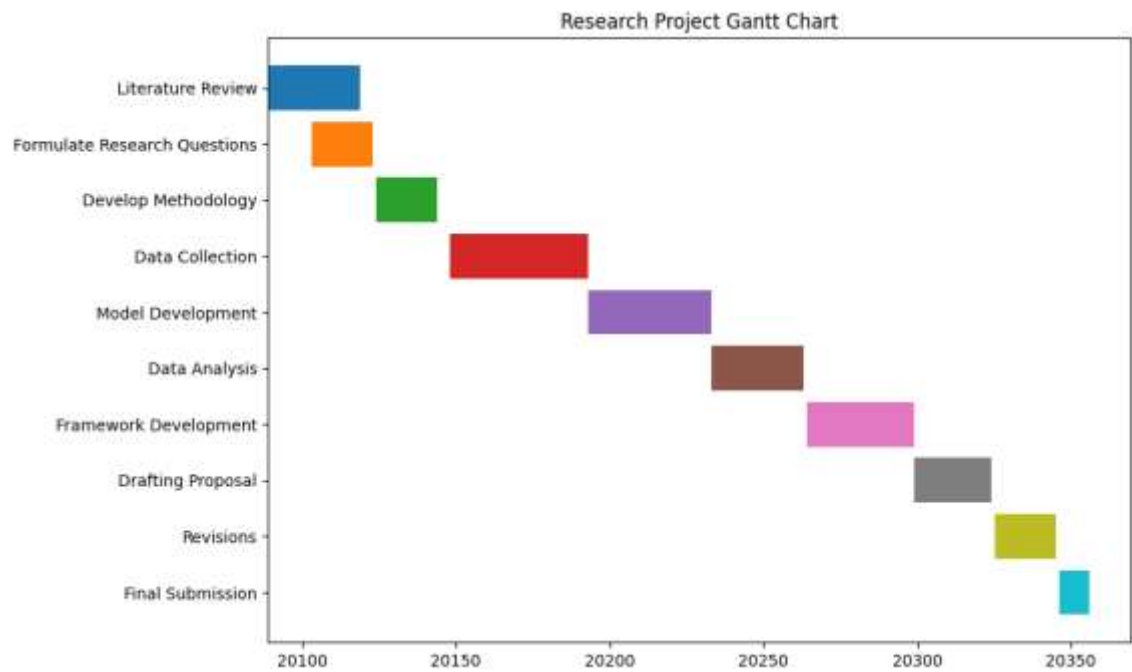
Finally, the findings are expected to support broader adoption of semantic search technologies in procurement. By reducing manual workload, improving matching accuracy, and streamlining decision making, the research has the potential to enhance efficiency across supply chain operations.

References

- Allahim, A., Cherif, W., & Imine, A. (2025). Semantic approaches for query expansion: Taxonomy, challenges, and future research directions. PeerJ.
- Khan, M., Abbas, S., & Binsawad, M. (2024). Secure semantic search using deep learning in a blockchain-assisted multi-user setting. Journal of Cloud Computing.
- Mahboub, M., Za'ter, A., Al-Rfooh, H., Estaitia, M., Jaljuli, A., & Hakouz, M. (2024). Evaluation of semantic search and its role in retrieval-augmented generation for Arabic. arXiv:2403.18350.
- Yadav, R., & Gopinathan, G. (2025). Semantic search and retrieval-augmented generation for academic literature analysis using ColBERT.
- Roy, S., Modak, R., Barik, R., & Goon, A. (2019). An overview of semantic search engines. International Journal of Research and Review.
- Foote, K. (2023). Semantic technology trends in 2024. Dataversity.
- Karpukhin, V., et al. (2020). Dense passage retrieval for open-domain question answering. ACL.
- Thakur, N., et al. (2021). BEIR: A heterogeneous benchmark for zero-shot evaluation of retrieval models. NeurIPS.
- Santhanam, K., et al. (2022). ColBERTv2: Effective and efficient retrieval via lightweight late interaction.
- Zhang, Y., & Zhao, L. (2023). Semantic embedding techniques for e-commerce product search. IEEE Access.

Gantt Chart

The following Gantt chart illustrates the planned timeline for conducting this research. It outlines each major phase including literature review, data preparation, model development, evaluation, and final documentation along with their respective durations. This structured timeline ensures that the project progresses methodically and within the allocated timeframe.



The table below provides additional details for each research activity, including start dates and estimated duration in days:

Task	Duration (Days)
Literature Review	30
Formulate Research Questions	20
Develop Methodology	20
Data Collection	45
Model Development	40
Data Analysis	30
Framework Development	35
Drafting Proposal	25
Revisions	20
Final Submission	10