Semantic Search in AI for Automated Matching

Problem Statement

As more companies move their procurement and inventory processes into digital systems, a persistent problem continues to show up: customers don't describe products in a consistent or predictable way. One customer might enter the full technical name of an item, another might use only part of the name or a brand label, and some give vague descriptions such as "blue cable" or "the usual filter." On top of that, spelling errors and missing details often make the text even harder to interpret. Traditional keyword-based search tools struggle with this kind of variation because they depend heavily on exact word matches. When the system can't figure out the intended item, an employee has to step in, which slows down the workflow and increases the chance of mistakes a problem that becomes more noticeable as the catalog grows.

Because of these limitations, organizations have turned toward AI-based semantic search as a possible solution. Instead of matching words directly, semantic search tries to interpret what the user means. Still, these approaches come with their own challenges. Many models don't handle messy or incomplete descriptions well, and they often struggle when terminology shifts or when the system must work across different languages. Some techniques require very large, specialized datasets that most companies simply don't have. Even methods that perform well in research environments sometimes fail to adjust to the pace and variety of real business settings. This disconnect highlights a gap between what the tools are supposed to do and what organizations actually need.

Since accurate item matching is central to automated procurement, this gap has practical consequences: delays, higher costs, and unnecessary manual corrections. Developing a more adaptable approach one that can handle different ways people describe items is essential. Before such a system can be created, it's important to understand what the existing methods do well, where they fall short, and what might realistically be improved in a business environment.

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.

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, embeddings 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, embeddings, 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.

References

1. Allahim, Cherif & Imine (2025). Semantic approaches for query expansion: taxonomy, challenges, and future research directions. Semantic approaches for query expansion: taxonomy, challenges, and future research directions [PeerJ]

2. Khan, Abbas & Binsawad (2024). Secure semantic search using deep learning in a blockchain-assisted multi-user setting. Secure semantic search using deep learning in a blockchain-assisted multi-user setting | Journal of Cloud Computing | Full Text

3. Mahboub, Za'ter, Al-Rfooh, Estaitia, Jaljuli & Hakouz (2024). Evaluation of semantic search and its role in retrieval-augmented generation for Arabic. [2403.18350] Evaluation of Semantic Search and its Role in Retrieved-Augmented-Generation (RAG) for Arabic Language

4. Foote (2023). Semantic technology trends in 2024. Semantic Technology Trends in 2024 - Dataversity