Mittuniversitetet

MID SWEDEN UNIVERSITY

Integrating Semantic with Keyword Product search in Online Marketplaces

By: Roberto Piran Amedi & Albin Rönnkvist

Overview

1. Introduction
2. Literature Review
3. Results and Discussion
4. Proposed Solution
5. Conclusions and Future Recommendations
6. Questions

Mittuniversitetet
MID SWEDEN UNIVERSITY

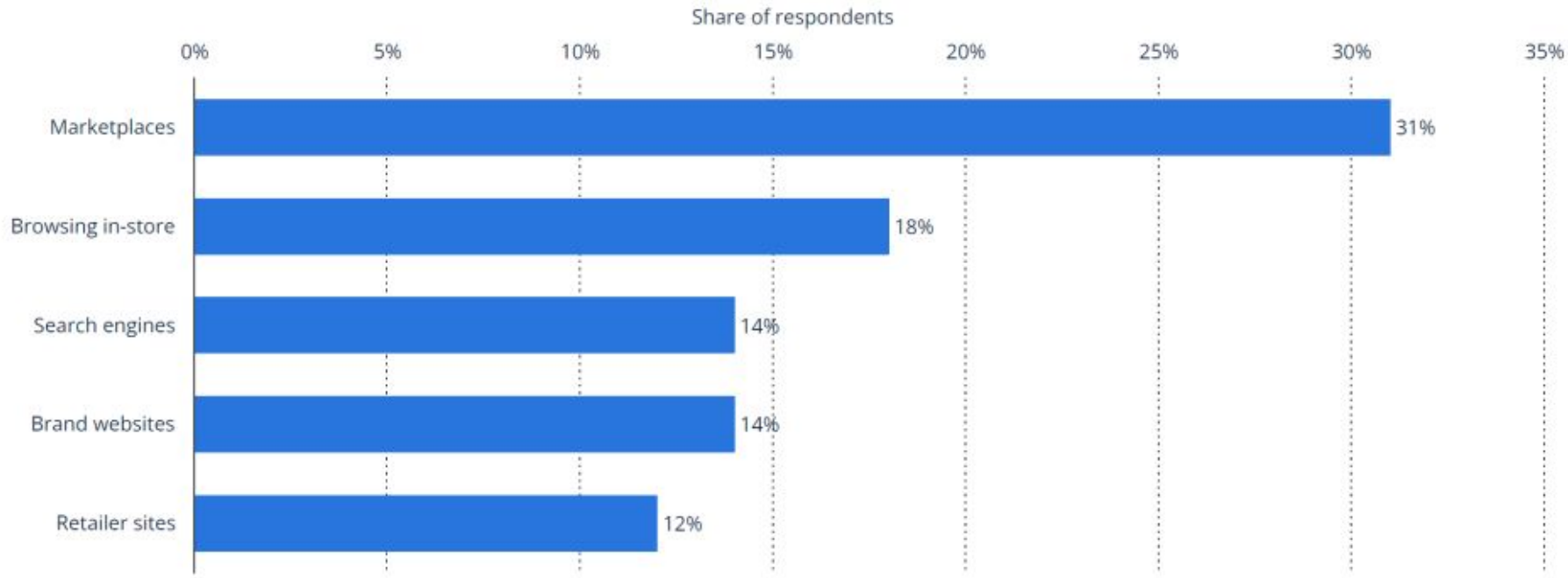
Introduction

Importance of online marketplaces and product search

Mittuniversitetet
MID SWEDEN UNIVERSITY

Introduction | Online Marketplaces

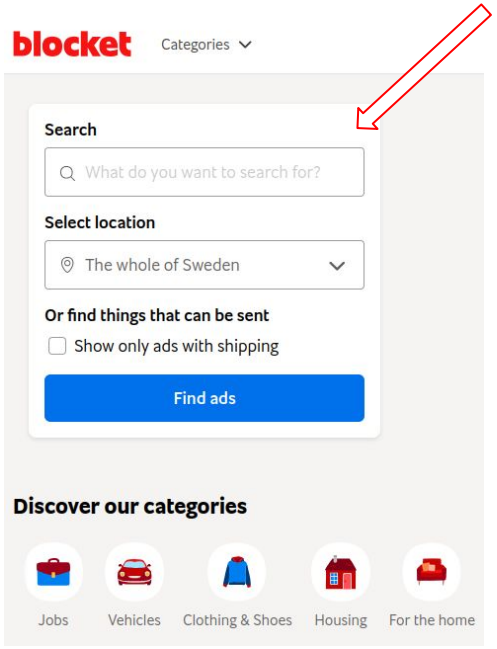
- Online marketplaces are popular platforms for finding products
- *2023 International Trade Administration Survey*: Online marketplaces are the main starting point for product searches worldwide

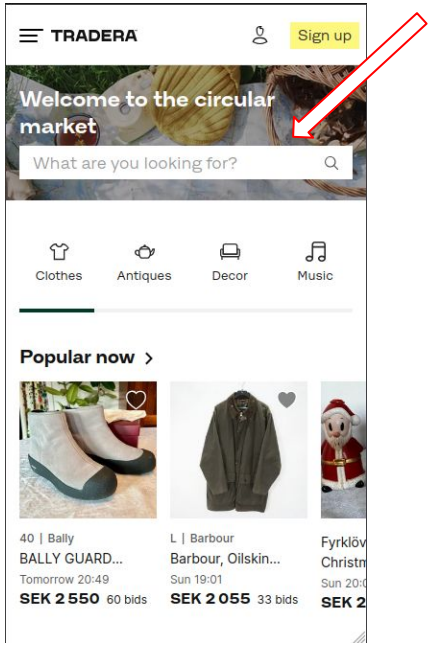


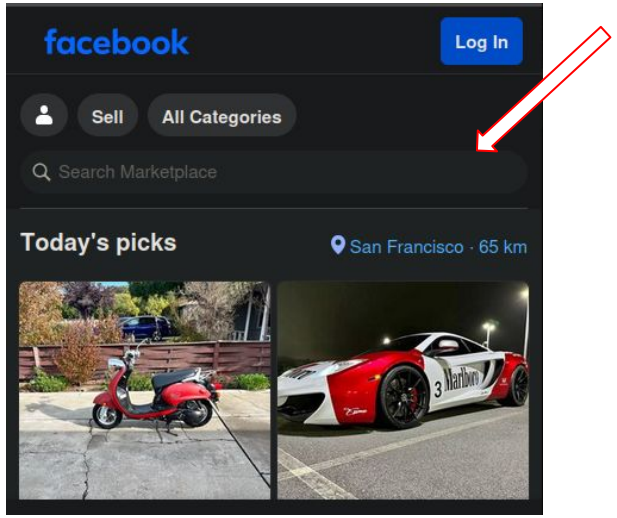
Mittuniversitetet
MID SWEDEN UNIVERSITY

Introduction | Product Search

- The search bar is a widely used tool for finding products
- *2022 Google Survey*: 7 out of 10 shoppers use the search bar as the primary method







Mittuniversitetet
MID SWEDEN UNIVERSITY

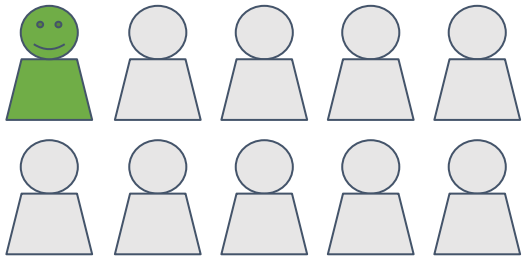
Introduction

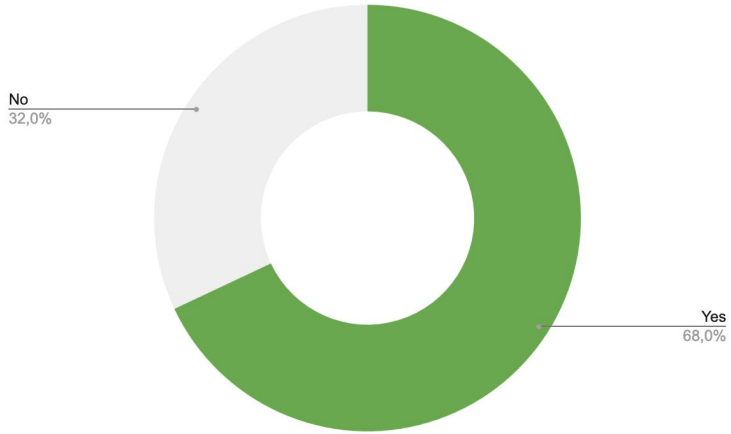
Current state and challenges of product search

Mittuniversitetet
MID SWEDEN UNIVERSITY

Introduction | Current State

- There is widespread dissatisfaction with search results among shoppers
- *2022 Google Survey*: Only 12% of shoppers find exactly what they are searching for when using the search bar
  - Results in $2 trillion in annual losses globally
- *2024 Constructor Survey*: 68% of shoppers want improved search results

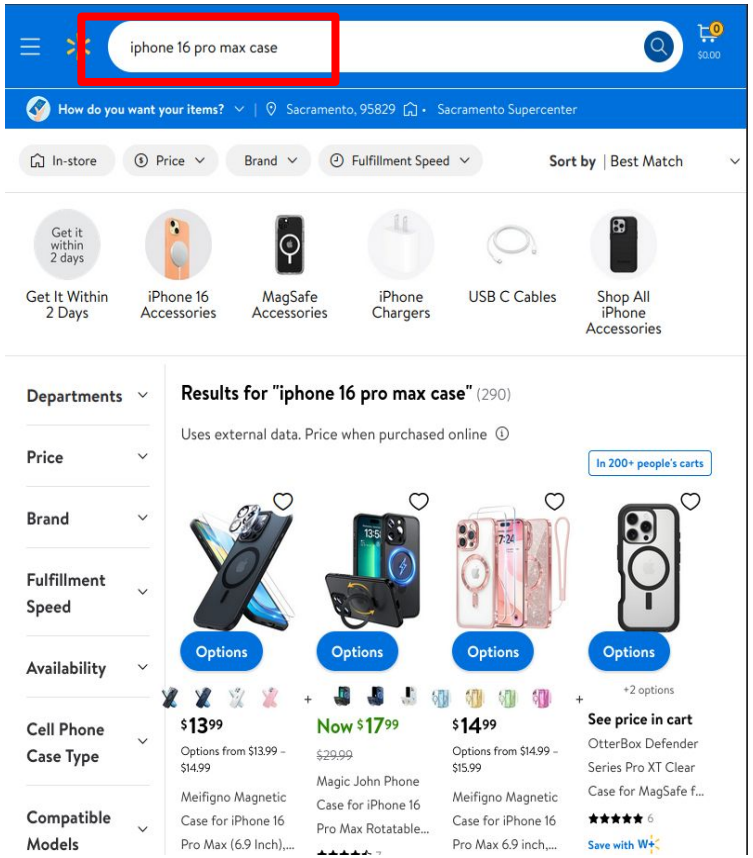


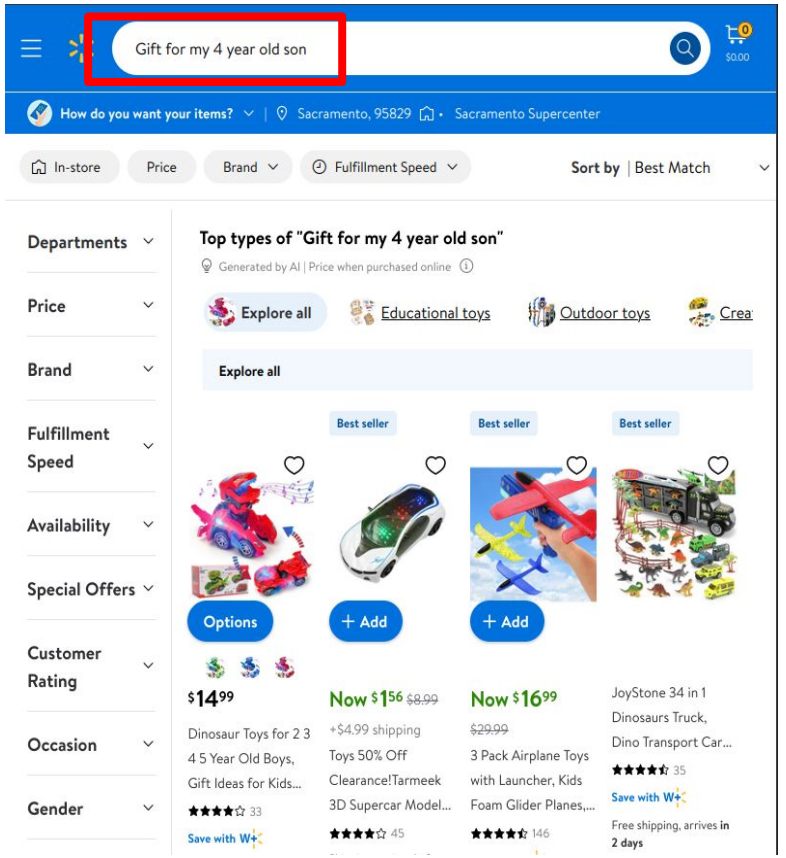


Mittuniversitetet
MID SWEDEN UNIVERSITY

Introduction | Challenges

- Diverse user intentions



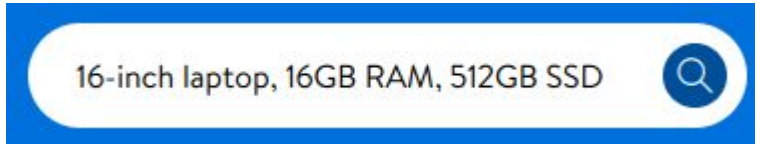


Mittuniversitetet
MID SWEDEN UNIVERSITY

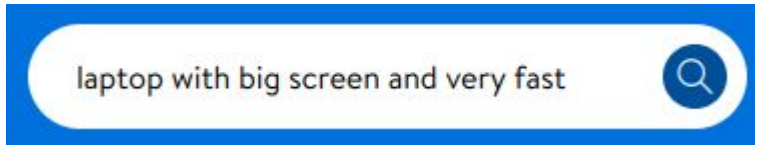
Introduction | Challenges

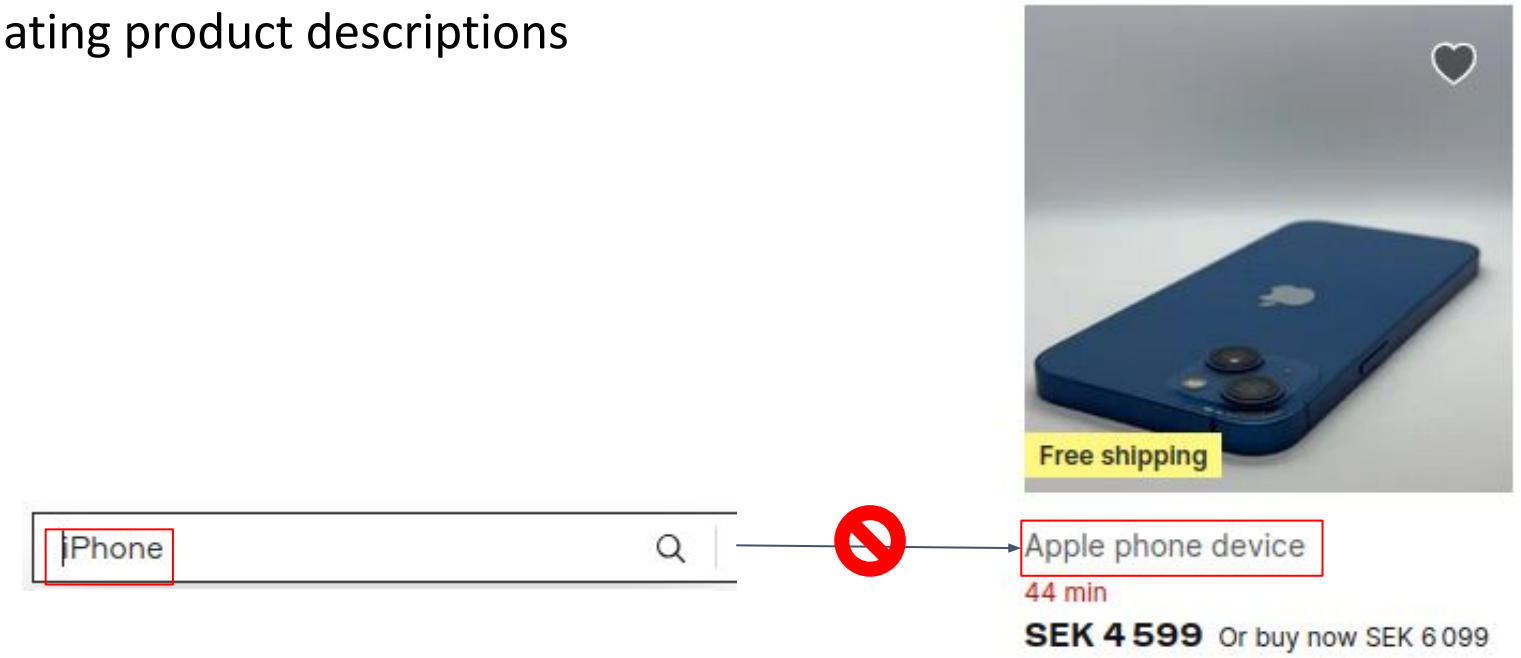
- Diverse user base
  - Language and culture
  - Backgrounds and expertise
- Formulating queries
- Formulating product descriptions

Experienced computer user



Inexperienced computer user





Mittuniversitetet
MID SWEDEN UNIVERSITY

Introduction

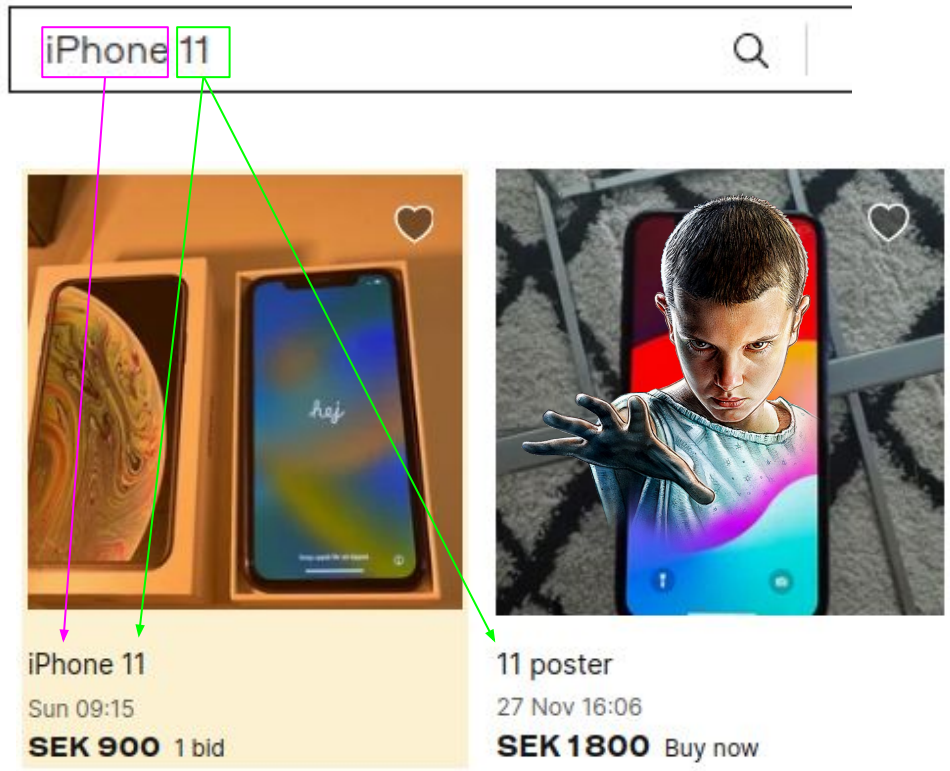
Common product search systems: keyword-based, semantic-based, and hybrid

Introduction | Keyword Search

- Word matching

- **Main strength**: Precise matching (e.g., model numbers and brand names)

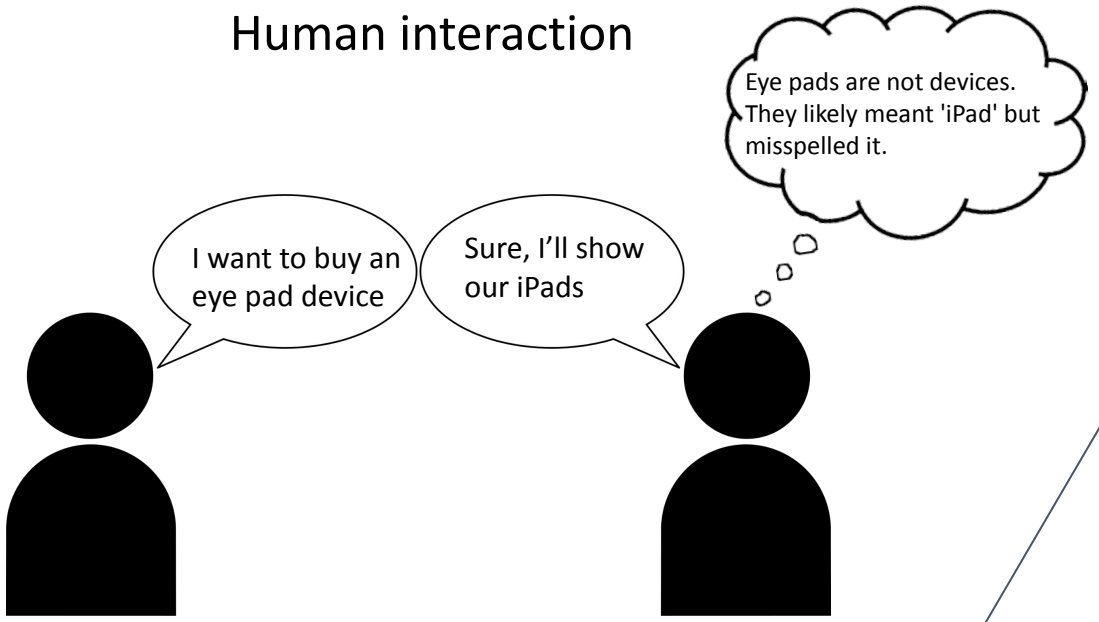
- **Limitations**:
  - Lacks understanding of context and underlying intent
  - Struggles with vocabulary gaps (e.g., misspellings and synonyms)



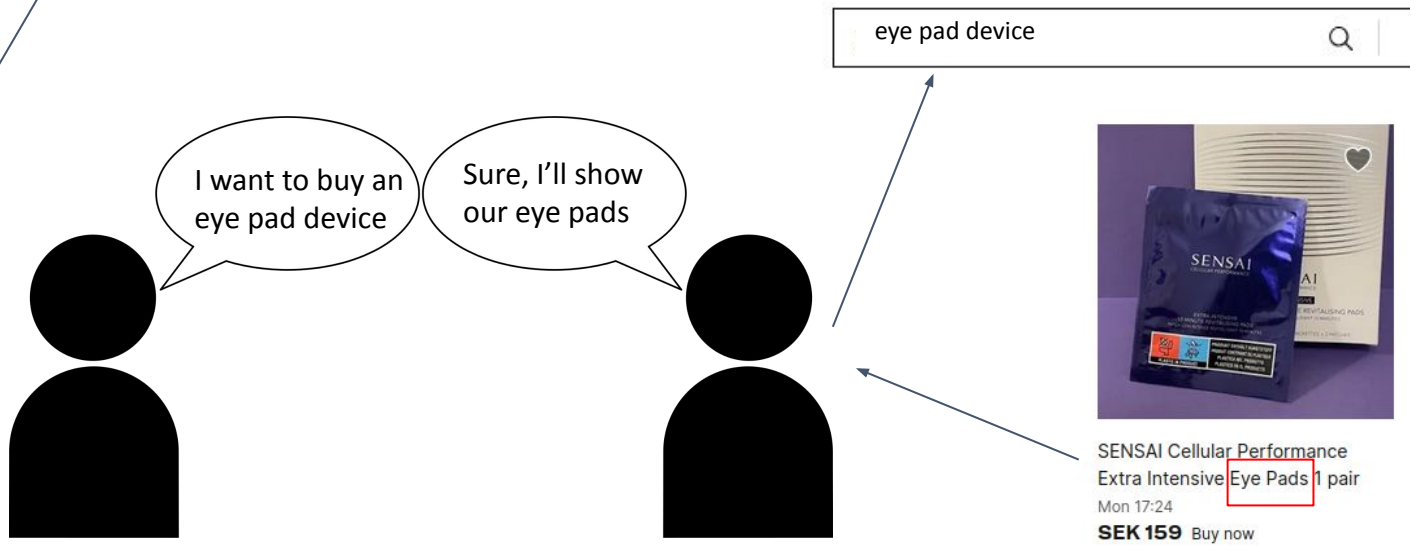
Mittuniversitetet
MID SWEDEN UNIVERSITY

Introduction | Keyword Search

Human interaction

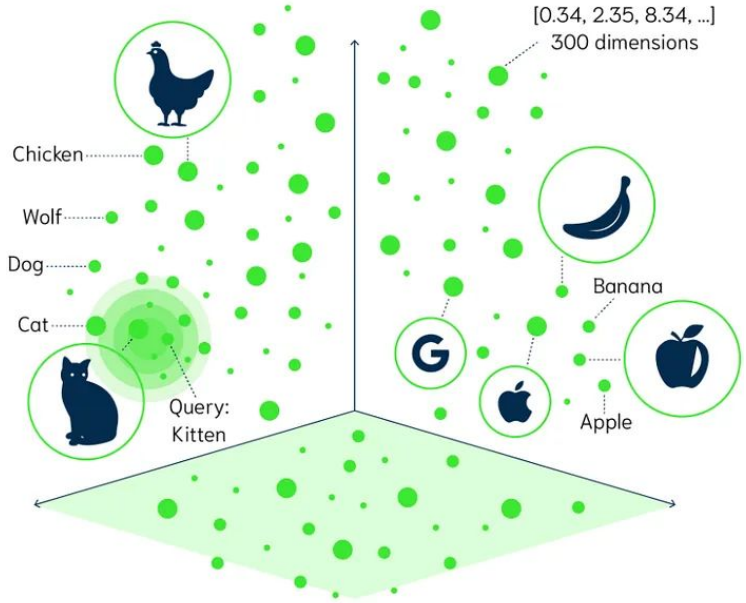


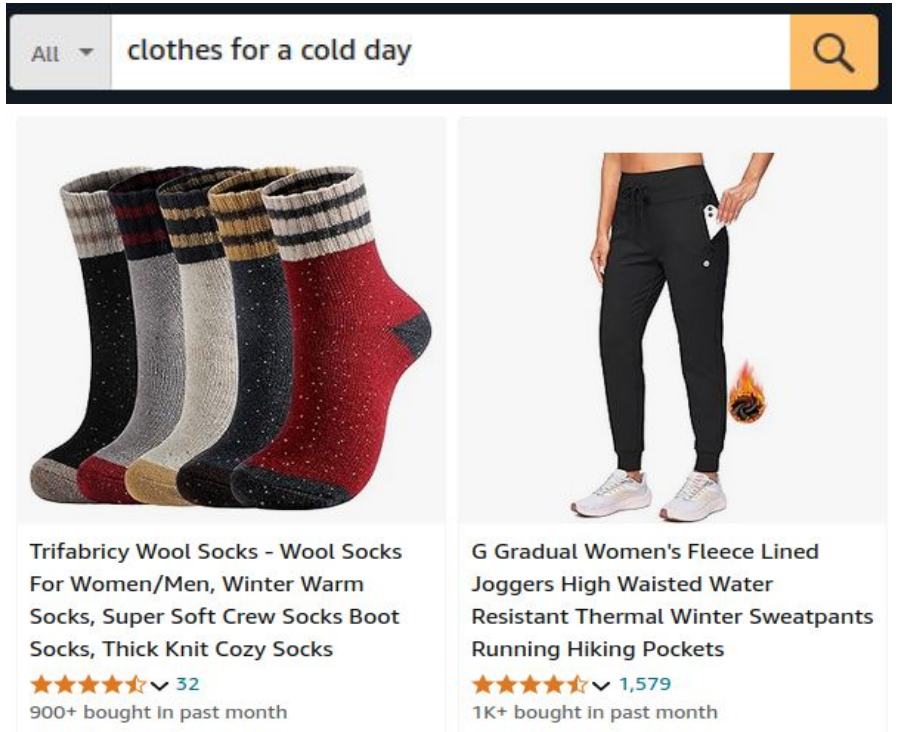
Keyword-based search system



Introduction | Semantic Search

- Matching based on the intent and contextual meaning
- “Product / query” -> Embedding model -> [0, 7, 2, …]
  - Captures semantic meaning
  - Measure distance or similarity
- **Main strength**: Understands the underlying meaning, beyond just words
- **Limitations**:
  - May produce inaccuracies when specific terms are important (e.g., model numbers)
  - Other known limitations, such as negations and units of measurement

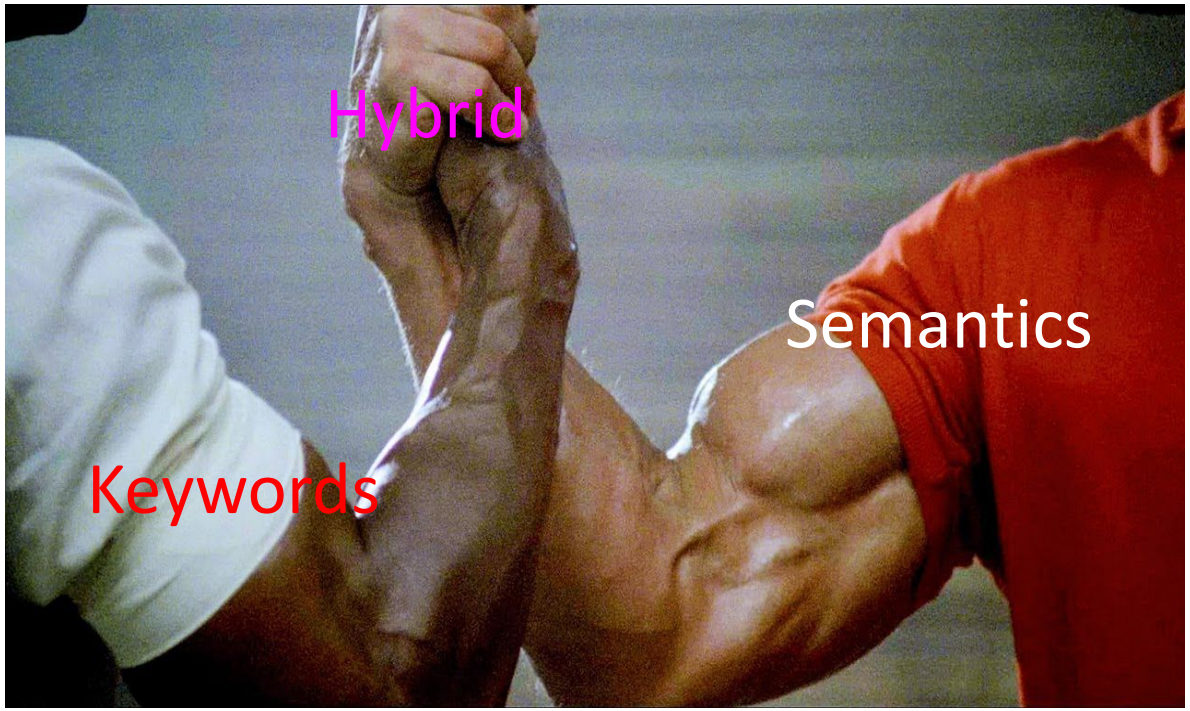




Introduction | Hybrid Search

- Combines keyword and semantic search

- Complementary:
  - Precise term matching
  - Underlying meaning



Mittuniversitetet
MID SWEDEN UNIVERSITY

Introduction

Research Questions

Mittuniversitetet
MID SWEDEN UNIVERSITY

Introduction | Research Questions

- **Hybrid Architectures:** How can hybrid models be implemented to utilize the capabilities of both keyword and semantic models?
- **Hybrid Search Impact on Product Search Results:** What is the impact of hybrid models on product search results compared to keyword and semantic models used independently?

Mittuniversitetet
MID SWEDEN UNIVERSITY

Literature Review

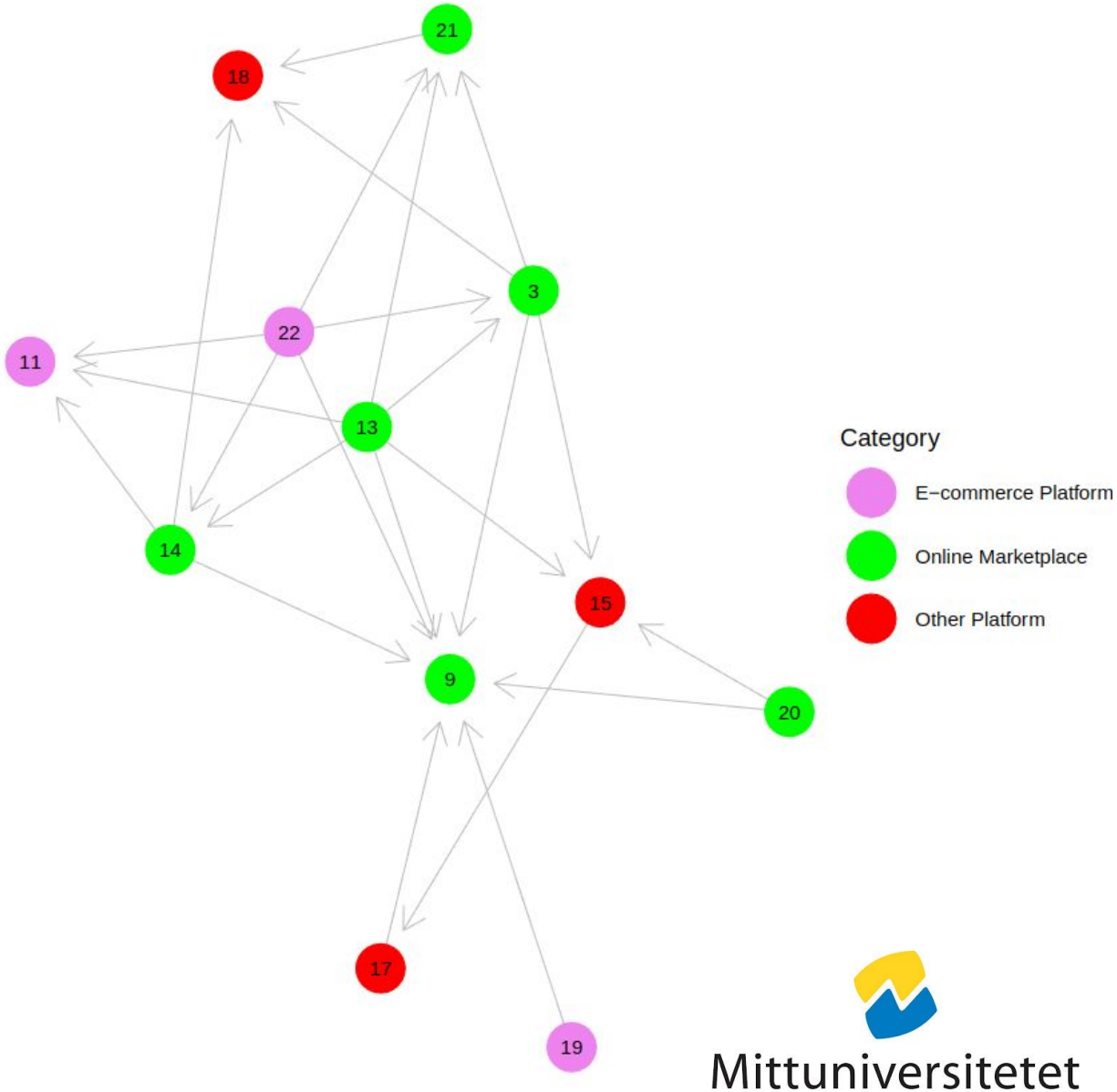
Selected studies, hybrid architectures and their impact on product search results

Mittuniversitetet

MID SWEDEN UNIVERSITY

Literature Review

- Selected studies:
  - 6 online marketplaces
  - 3 e-commerce platforms
  - 3 other platforms
- Includes:
  - Hybrid models
  - Offline and online experiments
- The studies are interconnected



Mittuniversitetet
MID SWEDEN UNIVERSITY

Results and Discussion

Hybrid architectures and their impact on product search results

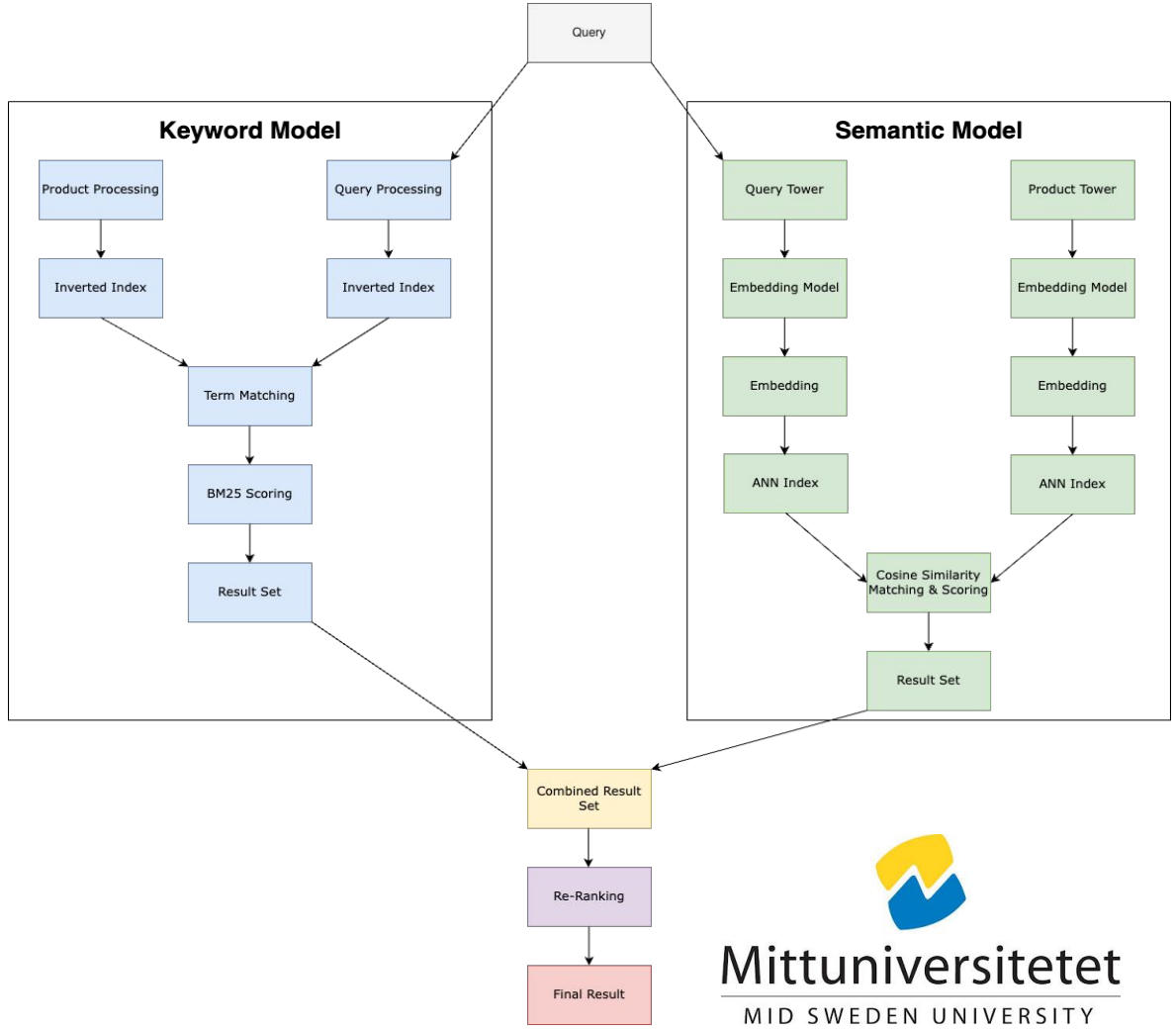
Results and Discussion | Hybrid Architectures

- Differences between hybrid models, but similar foundation

- **Core Keyword Models:**
  - Inverted index term matching
  - BM25 scoring

- **Integration of Semantic Models:**
  - Two-tower architecture with query and product embeddings
  - Approximate Nearest Neighbor (ANN) search with cosine similarity for matching and scoring

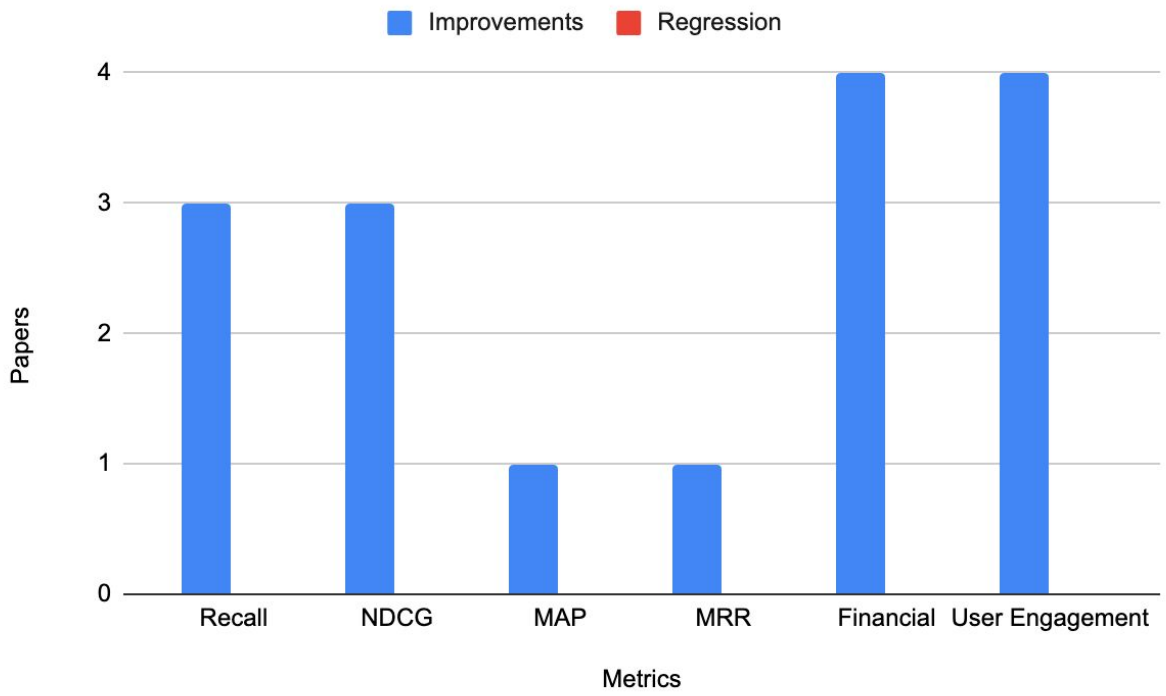
- **Hybridization:**
  - **Merging:** Combines results from both models into a unified match set
  - **Re-ranking:** Following this merging, refines and ranks matches based on relevance (limited re-ranking details available across studies)



Mittuniversitetet
MID SWEDEN UNIVERSITY

Results and Discussion | Impact

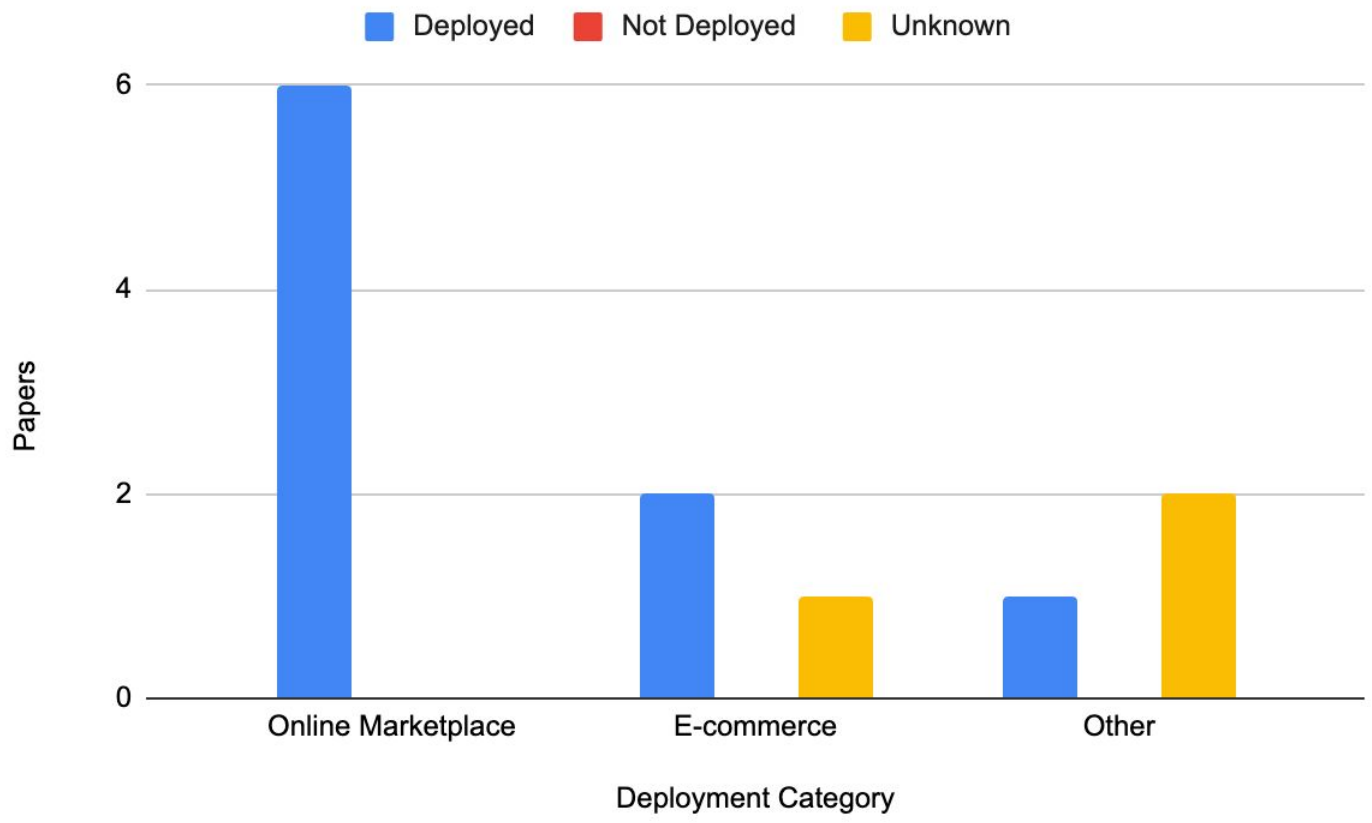
- Significant improvements in product search result metrics when compared to standalone models
- Improved coverage, ranking quality and business performance



Mittuniversitetet
MID SWEDEN UNIVERSITY

Results and Discussion | Impact

- A majority of hybrid models were deployed in production

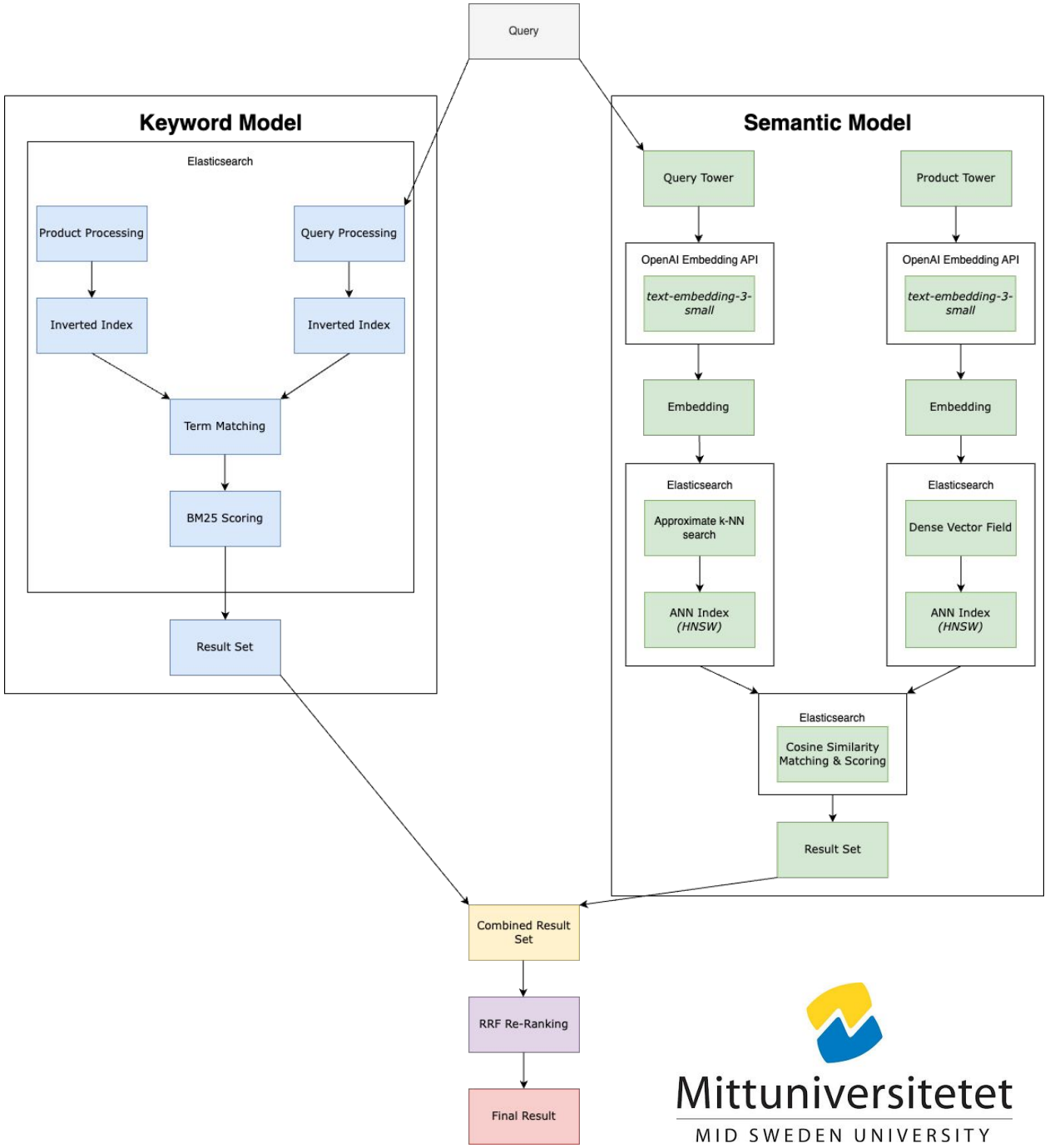


Mittuniversitetet
MID SWEDEN UNIVERSITY

Proposed Solution

Proposed Solution

- Hybrid models provide better product search results
- Hybrid models can be complex to implement
  - Requires expertise and resources
- **Proposed solution:** A simplified hybrid model for broader adoption
  - Inspired by models in the review
  - Lightweight and flexible
  - [Codebase]



Mittuniversitetet
MID SWEDEN UNIVERSITY

Conclusions and Future Recommendations

Conclusions and Future Recommendations

- Discussed different hybrid architectures and identified a similar foundation across the literature
- Multiple studies, with various experiments and metrics, verified that hybrid models significantly enhance product search results
- We proposed a simplified hybrid model as a starting point
- Suggest further exploration of the overall hybrid model impact:
  - Complexity and maintenance
  - Resource requirements
  - Cost implications
  - Latency
  - Pagination

Mittuniversitetet

MID SWEDEN UNIVERSITY

Questions