

# A Comparison of Classical and Modern Information Retrieval Approaches on Recipes

<https://github.com/Dowakiin/Advanced-Information-Retrival-WS24-25-Group22>

A Project by Group 22:

- Markus Auer-Jammerbund
- Thomas Knoll
- Jonas Pfisterer
- Thomas Puchleitner

Team-role: Query pipeline setup, Evaluation, Report

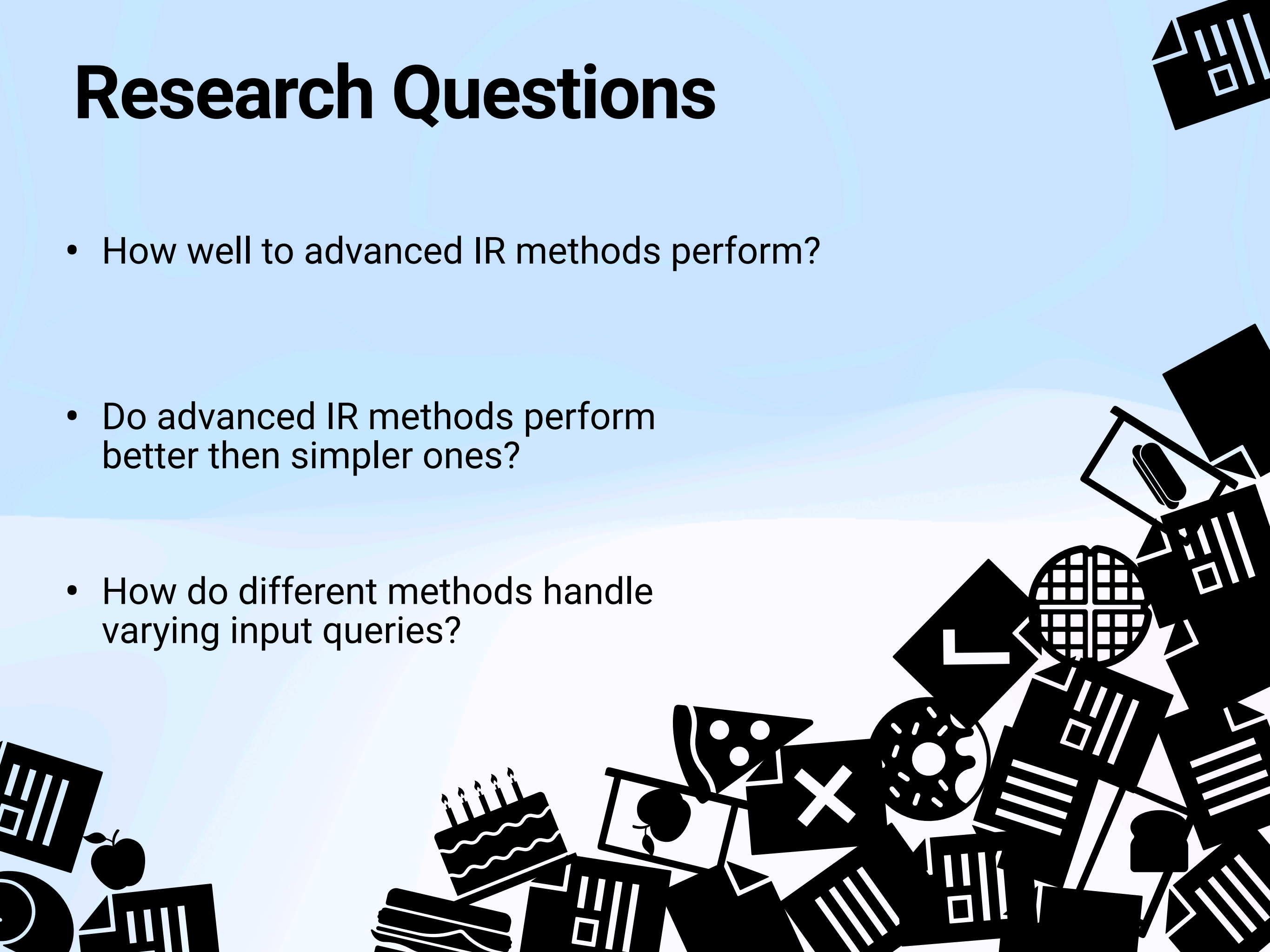
Team-role: Design Document, *BERT* embeddings, Evaluation, Report

Team-role: *Word2Vec* embeddings, Evaluation, Report

Team-role: Design Document, *TF-IDF* embeddings, Result processing, Evaluation, Report

# Research Questions

- How well to advanced IR methods perform?
- Do advanced IR methods perform better then simpler ones?
- How do different methods handle varying input queries?



# Data and Models

Where did we get it from and what did we do with it?

## Data:

- From Hugging Face

## Pre-processing:

- Merged the ingredient and instruction columns
- Reduced the size to 100,000 (from 2 mil.)

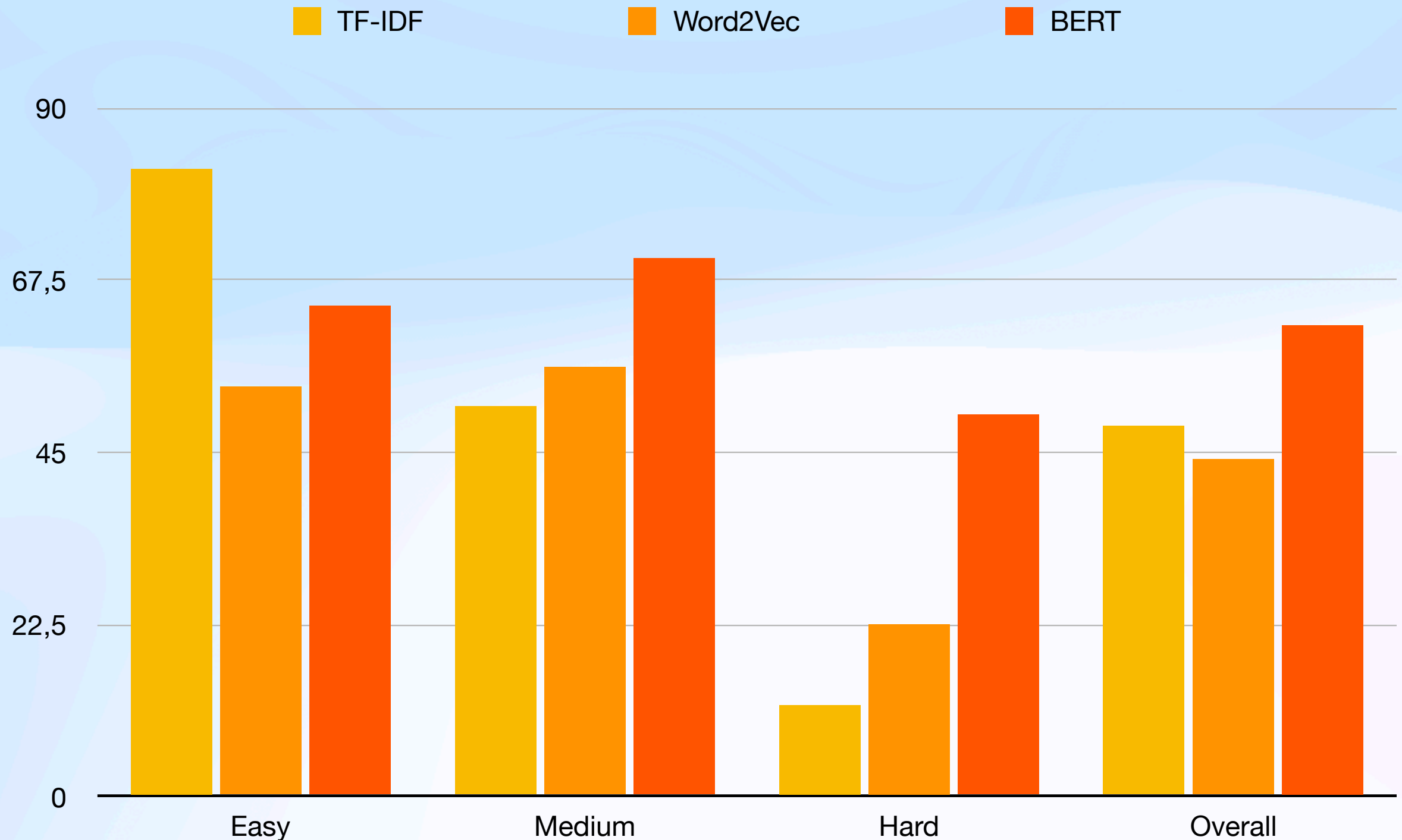


## Methods:

- TF-IDF: The old reliable
- Word2Vec: A newer approach
- BERT: State-of-the-art heavyweight

# Results

## Recipe relevance per IR method for different query types



# Conclusion / Possible future work

- Test with the whole dataset
- Increase robustness by: asking more queries, utilizing additional evaluators, and expanding on the returned recipes
- Deeper analysis of the weak performance of *Word2Vec* (focusing on its parameter settings, training conditions, etc.)
- Additional training or fine-tuning of the BERT model