

# Comparative Error Analysis of Sparse and Dense Retrieval Models

## TF-IDF:

Not in top 5: 88/1000

Not in top 2: 157/100

Not in ranks: 57/100

TF-IDF MRR: 0.7160953661663415

TF-IDF Accuracy: 0.616

Q45: What type of fossils were found in China?

Top: Because of their soft, gelatinous bodies, ctenophores are extremely rare as fossils, and fossils that have been interpreted as ctenophores have been found only in lagerstätten, places where the environment was exceptionally suited to preservation of soft tissue. Until the mid-1990s only two specimens good enough for analysis were known, both members of the crown group, from the early Devonian (Emsian) period. Three additional putative species were then found in the Burgess Shale and other Canadian rocks of similar age, about 505 million years ago in the mid-Cambrian period. All three apparently lacked tentacles but had between 24 and 80 comb rows, far more than the 8 typical of living species. They also appear to have had internal organ-like structures unlike anything found in living ctenophores. One of the fossil species first reported in 1996 had a large mouth, apparently surrounded by a folded edge that may have been muscular. Evidence from China a year later suggests that such ctenophores were widespread in the Cambrian, but perhaps very different from modern species – for example one fossil’s comb-rows were mounted on prominent vanes. The Ediacaran *Eoandromeda* could putatively represent a comb jelly.

Gold: The early Cambrian sessile frond-like fossil *Stromatoveris*, from China’s Chengjiang lagerstätte and dated to about 515 million years ago, is very similar to *Vendobionta* of the preceding Ediacaran period. De-Gan Shu, Simon Conway Morris et al. found on its branches what they considered rows of cilia, used for filter feeding. They suggested that *Stromatoveris* was an evolutionary “aunt” of ctenophores, and that ctenophores originated from sessile animals whose descendants became swimmers and changed the cilia from a feeding mechanism to a propulsion system.

Q17: What are one of the key cell types of the adaptive immune system?

Top: Dendritic cells (DC) are phagocytes in tissues that are in contact with the external environment; therefore, they are located mainly in the skin, nose, lungs, stomach, and intestines. They are named for their resemblance to neuronal dendrites, as both have many spine-like projections, but dendritic cells are in no way connected to the nervous system. Dendritic cells serve as a link between the bodily tissues and the innate and adaptive immune systems, as they present antigens to T cells, one of the key cell types of the adaptive

immune system.  
Gold at rank 0!

Q292: Where does the Rhine empty?

Top: In the centre of Basel, the first major city in the course of the stream, is located the "Rhine knee"; this is a major bend, where the overall direction of the Rhine changes from West to North. Here the High Rhine ends. Legally, the Central Bridge is the boundary between High and Upper Rhine. The river now flows North as Upper Rhine through the Upper Rhine Plain, which is about 300 km long and up to 40 km wide. The most important tributaries in this area are the Ill below of Strasbourg, the Neckar in Mannheim and the Main across from Mainz. In Mainz, the Rhine leaves the Upper Rhine Valley and flows through the Mainz Basin.

Rank: No match

Gold: The Rhine (Romansh: Rein, German: Rhein, French: le Rhin, Dutch: Rijn) is a European river that begins in the Swiss canton of Graubünden in the southeastern Swiss Alps, forms part of the Swiss-Austrian, Swiss-Liechtenstein border, Swiss-German and then the Franco-German border, then flows through the Rhineland and eventually empties into the North Sea in the Netherlands. The biggest city on the river Rhine is Cologne, Germany with a population of more than 1,050,000 people. It is the second-longest river in Central and Western Europe (after the Danube), at about 1,230 km (760 mi), with an average discharge of about 2,900 m<sup>3</sup>/s (100,000 cu ft/s).

**BM25:**

Not in top 5: 58/1000

Not in top 2: 97/1000

Not in ranks: 43/1000

BM25 MRR: 0.7944127703565786

BM25 Accuracy: 0.717

Q45: Gold at Rank 18. Top prediction same as TF-IDF.

Q292: Where does the Rhine empty?

Top: Same as TF-IDF

Rank: No match

Q17: Gold at rank 0!

**DPR:**

Not in top 5: 126/1000

Not in top 2: 237/1000

Not in ranks: 140/1000

DPR MRR: 0.5511221684533213

DPR Accuracy: 0.434

Q45: Gold at Rank 2! Top prediction about geology/fossils in asian countries.

Q292: Gold at Rank 0!

Q17: No exact match

Top: The adaptive immune system evolved in early vertebrates and allows for a stronger immune response as well as immunological memory, where each pathogen is "remembered" by a signature antigen. The adaptive immune response is antigen-specific and requires the recognition of specific "non-self" antigens during a process called antigen presentation. Antigen specificity allows for the generation of responses that are tailored to specific pathogens or pathogen-infected cells. The ability to mount these tailored responses is maintained in the body by "memory cells". Should a pathogen infect the body more than once, these specific memory cells are used to quickly eliminate it.

BM25 generally performs better than TF-IDF, especially with reference to the number of exact/close matches, as evidenced by the better accuracy rate. It also has the gold passages in its rankings even in cases where TF-IDF does not include the gold passage in top-20. BM-25 is obviously an improvement over TF-IDF as it accounts for more factors such as document length.

DPR model, unlike BM25 and TF-IDF, is trained as opposed to being directly computed through document/term frequencies/lengths. Hence, it can account for contexts and semantics - factors that are quite important but beyond the scope of previous models. As the model is not fine-tuned, it doesn't perform quite as well, but the importance of including semantics can still be evidenced. In both Q45 and Q292, DPR outperforms previous models whose outputs are naive attempts at maximising key-word matching, while more context is required to predict the gold passage to correctly answer the question, even if the key-word is only used once. For Q45, DPR's top prediction is a passage regarding geology/fossils in asian countries including China, which seems like a reasonable guess especially when the gold passage itself is not so obvious. It is the same case for Q292, where the previous models fail to include gold passage in the ranking, but DPR correctly predicts it at rank 0.

For Q17, DPR however does not have an exact match in its rankings, while BM25 has the answer at rank 0 and TF-IDF also has the answer, although at rank 7. This is one case where the maximising the key-words with respect to other frequencies worked out in the models' favour and the context was too ambiguous or similar to other potential passages that the DPR model did not include the answer in its rankings. In such cases, DPR does not assign enough weight to frequency/phrase based features and lose out to simpler sparse retrievers. Ensemble methods or neural models capable of imitating sparse retrievers in particular cases would work better in both cases. Such hybrid models would have both the expressiveness of dense neural models along with the precision and simplicity of sparse methods.