Information Retrieval Evaluation: Precision & Recall Analysis

Introduction

This report evaluates an Information Retrieval (IR) system's performance based on **Precision** and **Recall** using Boolean retrieval. The evaluation is conducted on three queries:

- **Query 1:** "swim and water"
- **Query 2:** "dinner or eating"
- **Query 3:** "winter or snow"
- **Query 4:** "children not dangerous"

We will demonstrate the system's performance with these queries and analyze the results.

Methodology

Evaluation Metrics

- **Precision** measures the proportion of retrieved documents that are relevant.

- **Recall** measures the proportion of relevant documents that were retrieved.

Query 1: "swim and water"

Retrieved Documents

- **Found Documents:** `RubihornTrail_en.txt`, `KaiserEgg_en.txt`, `Murgsee_en.txt`, `Schafsiedel_en.txt`

Relevant Documents

- **Expected Relevant:** `RubihornTrail_en.txt`, `Murgsee_en.txt`, `Schafsiedel_en.txt`

Irrelevant Documents

- `KaiserEgg_en.txt` (irrelevant due to contradictory statement: "So I wouldn't necessarily swim in this standing water.").

Missing Documents

- `Gruensee_en.txt`, `RoterGrat_en.txt` (expected but not retrieved).

Precision & Recall

- **Precision:**
  Precision = 3 / 4 = 0.75

- **Recall:**
  Recall = 3 / 5 = 0.6

Query 2: "dinner or eating"

Retrieved Documents

- **Found Documents:** `ZugspitzeTrail2_en.txt`, `Gatterl_en.txt`, `Rossmad_en.txt`, `ZugspitzeTrail1_en.txt`

Relevant Documents

- All retrieved documents are relevant.

Missing Documents

- `Mutoerl_en.txt`, `Wandflue_en.txt` (expected but not retrieved).

Precision & Recall

- **Precision:**
  Precision = 4 / 4 = 1.0

- **Recall:**
  Recall = 4 / 6 = 0.67

Query 3: "winter or snow"

Retrieved Documents

- **Found Documents:** `Juifen_en.txt`, `Wandflue_en.txt`, `Pochtenfall_en.txt`, `Olpererhuette_en.txt`, `Schafsiedel_en.txt`, `Sonnjoch_en.txt`, `Hoellental_en.txt`, `ZugspitzeTrail1_en.txt`, `Aurlandsvangen_en.txt`, `Puehringerhuette_en.txt`, `Oeschinensee_en.txt`, `Sonntagshorn_en.txt`, `Lackenkogel_en.txt`, `Rinnenspitze_en.txt`, `Geissspitze_en.txt`

Irrelevant Documents

- `Pochtenfall_en.txt`, `Hoellental_en.txt`, `Oeschinensee_en.txt`, `Lackenkogel_en.txt`.(irrelevant due to contradictory statements).

Missing Documents

- `Tseuzier_en.txt`, `Cimetta_en.txt` (expected but not retrieved, even though they contain information about ski (relationed to winter and snow)).

Precision & Recall

- **Precision:**
  Precision = 10 / 15 = 0.67

- **Recall:**
  Recall = 10 / 12 = 0.83

Query 4: "children not dangerous"

Retrieved Documents

- **Found Documents:** `Gratlspitze_en.txt`, `CreuxDuVan_en.txt`, `Madrisella_en.txt`, `Grossarl_en.txt`

Relevant Documents

- All retrieved documents are relevant.

Missing Documents

- `Griessenkareck_en.txt` (expected but not retrieved because it uses the term "kids" instead of "children").

Precision & Recall

- **Precision:**
  Precision = 4 / 4 = 1.0

- **Recall:**
  Recall = 4 / 5 = 0.8

Discussion

Observations

- **Query 1:** Precision (0.75) is decent, but Recall (0.6) could be improved by retrieving missing relevant documents.
- **Query 2:** Perfect Precision (1.0), but Recall (0.67) is lower due to missing relevant documents.
- **Query 3:** High Recall (0.83), but Precision (0.67) is affected by irrelevant results.
- **Query 4:** The system retrieved the correct documents with perfect precision (1.0), but recall could be improved by capturing documents like `Griessenkareck_en.txt`, which used the term "kids" instead of "children".

Analysis and Conclusion

The Boolean model fulfills its functionality by returning the documents explicitly matching the query. However, the sometimes suboptimal Precision and Recall values indicate that this strict retrieval approach lacks contextual awareness. As a result, it retrieves some irrelevant documents while failing to return all relevant ones.

A system like the Vector Space Model (VSM) offers an advantage by incorporating ranking, which helps prioritize more relevant documents. An even more advanced approach would integrate AI-driven techniques to enhance retrieval effectiveness. For example:

Synonym Recognition: Automatically identifying and including synonyms in searches (e.g., recognizing "kids" as equivalent to "children").

Context Awareness: Understanding the surrounding words to disambiguate meanings (e.g., distinguishing "Apple" as a company from "apple" as a fruit when searching for stock-related information).

Implementing these improvements would significantly enhance both Precision and Recall, making the IR system more robust and context-aware.

Source Code & Dataset

The code and dataset used in this experiment are available on GitHub

- **[GitHub Repository Link]** (https://github.com/Tim10022023/InformationRetrievalSystem)