= 0.5625

= 56.25%

Q.18 In IR, if q is the information request and a set of relevant documents for query q is Rq = {d3, d5, d9, d25, d39, d44, d50, d70, d80, d120}. Consider new retrieval algorithm has been designed and has been evaluated for information request q returns, ranking of the documents in the answer set is as follows

1) d120
2) d84
3) d50
4) d6
5) d8
6) d9
7) d58
8) d129
9) d143
10) d25
11) d38
12) d48
13) d230
14) d113
15) d3

The documents that are relevant to the query are underlined. Calculate precision and recall for the documents that are relevant for query q.

Soln: Step 1: Given Data

. Relevant Documents set:

Rq = {d3, d5, d9, d25, d39, d44, d50, d70, d80, d120}

. Retrieved Ranking (is documents in order)

1. d120
2. d84
3. d50
4. d6
5. d8
6. d9
7. d58
8. d129
9. d143
10. d25
11. d38
12. d48
13. d230
14. d113
15. d3

Step 2: Find relevant documents retrieved from the ranking, the relevant ones are:

{d120, d50, d9, d25, d33}

Step 3: Precision

$$P = \frac{\text{Relevant Retrieved}}{\text{Total Retrieved}}$$

$$= \frac{5}{15}$$

$$= 0.3333$$

Step 4: Recall

$$R = \frac{5}{10} = 0.5$$

using logical operators (AND, OR, NOT) in combination with visual selection.
Example: Selecting customers where "Region = Asia ~~and~~ AND sales > 5000" using checkboxes and filters.

Q4. Explain user-oriented measures in performance evaluation of IR systems.

Ans. Traditional measures like Precision and Recall focus on system effectiveness but they do not fully reflect how useful results are from a users point of view. Hence user-oriented measures are used which consider remaining ranking graded relevance and user satisfaction.

1. Mean Reciprocal Rank
- Measures how quickly the system returns the first relevant document.
- Reciprocal Rank = $\frac{1}{\text{rank of first relevant document}}$
- MRR = Average of reciprocal ranks across all queries
- Higher MRR = user finds relevant result earlier.
- Normalized Discounted Cumulative Gain

Gain
Considers graded relevance
Rewards systems that place highly relevant

documents at the top ranks.
- Normalized between 0 and 1, where 1 = perfect ranking.

3. Mean Average Precision
- Average of Precision values computed at every position where a relevant document occurs.
- MAP = mean of AP across multiple queries

4. Cumulative Gain and Discounted Cumulative Gain
- CG = Sum of relevance scores of retrieved documents.
- DCG = discounted version of CG.

5. E - Measure
- A weighted measure combining precision and recall.
- Adjusts importance depending on whether user values Recall or Precision.

6. User Satisfaction Measures
- Based on user values feedback such as
  - Time taken to find required information
  - Number of queries needed
  - Overall Satisfaction with retrieved results.

Q.5 What is Relevance Judgement? Explain term group relevance judgements and pseudo relevance feedback.

ns. Relevance Judgement is the process of

Q.11 What are alternative measures used to evaluate system performance in IR

Ans 1) F-Measure / E-measure

Harmonic mean of precision and recall

$$F1 = \frac{2 \cdot P \cdot R}{P + R}$$

2) MRR - Mean Reciprocal Rank

3) Normalized Discounted Cumulative Gain (NDCG)

4) Specifity

1) F-Measure / F1-score

- The F-measure (or F1-score) is the harmonic mean of precision and recall.
- Formula: $F1 = \frac{2 \cdot P \cdot R}{P + R}$

where:

- P = precision = $\frac{TP}{TP + FP}$
- R = Recall = $\frac{TP}{TP + FN}$

It balances precision (how many predicted positives are actually correct) and recall (how many actual positives were retrieved).

Useful when you need a single score that considers both precision and recall.

2) MRR (Mean Reciprocal Rank)

- Reciprocal Rank = $\frac{1}{\text{rank of first relevant result}}$
- MRR is the average of reciprocal ranks over a set of queries.

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$$

where $rank_i$ is position of first relevant document for query$_i$.

Used in search engines and recommendation systems to measure ranking quality.

3) NDCG (Normalized Discounted Cumulative Gain)

DCG measures ranking quality by giving higher scores for relevant documents appearing earlier in the list.

$$DCG_p = \sum_{i=1}^{p} \frac{2^{rel_i} + 1}{\log_2(i+1)}$$

where rel_i is relevance score of result at position i.

NDCG is DCG normalized by ideal DCG (IDCG), so values range from 0 to 1.

$$NDCG = \frac{DCG}{IDCG}$$

Used in ranking problems (e.g., search engines, recommender systems).

4) Specificity

specifity measures how well the model identifies true negatives.

$$Specifity = \frac{TN}{TN + FP}$$

where:

- TN = True Negatives (correctly predicted negatives)
- FP = False Positives (incorrectly predicted positives)

It answers: of all the negative cases, how many did the model correctly classify as negative?

Important in medical testing (e.g., avoiding false alarms).

for search

Step 6: Compute NDCG

$$NDCG = \frac{DCG_5}{IDCG_5}$$

$$= \frac{6.149}{6.324}$$

$$\approx 0.972$$

Q3. What are various techniques used to specify query in information visualization?

Ans. Techniques for Query Specification in Information Visualization

1. Form-Based Queries
   - Users fill out forms with fields and conditions.
   - Common in database front-ends or dashboards.
   - Example: Selecting "Age > 30" and "Country = India" in a form before visualizing results.
2. Direct Manipulation (Interactive Filtering)
   - Users directly manipulate visual elements (drag, click, brush) to specify queries.
   - Example:
     - Dynamic Queries (e.g., sliders to filter by time or price).
     - Brushing and Linking: selecting data points in one visualization highlights related points in others.
3. Range and Point Selection
   - Selecting ranges or single points in visualization to filter data.
   - Example:
     - Dragging a rectangle on a scatterplot to query points inside it.
     - Selecting a time window in a timeline.
4. Query by Example (QBE)

users provide examples of data they are interested in, and the system retrieves similar patterns.

Example: In a stock trend visualization, drawing a "u-shape" curve to find stocks with similar patterns.

5. Boolean Queries

Using logical operators (AND, OR, NOT) in combination with visual selection.

Example: selecting customers where "Region = Asia ~~and~~ AND sales > 5000" using checkboxes and filters.

Explain user-oriented measures in perfor evaluation of IR systems.

Traditional measures like Precision and Re focus on system effectiveness but they n not fully reflect how useful results are from a users point of view. Hence user-ori measures are used which consider remainin ranking graded relevance and user satisfa

b. User satisfaction Measures
- Based on user values feedback such as
  - Time taken to find required information
  - Number of queries needed
  - Overall satisfaction with retrieved results.

Q.5 What is Relevance Judgement? Explain term group relevance judgements and pseudo relevance feedback

Ans. Relevance Judgement is the process of deciding whether a document retrieved by an information Retrieval system is relevant to the user's query or not.
- It is usually performed by humans assessors or users, who examine documents and assign a relevance score.
- These judgement are the ground truth

for evaluating IR systems using measures like Precision, Recall, NDCG, etc.

Types of Relevance Judgements

a. Binary relevance: Document is either relevant or not relevant

b. Graded relevance: Documents are assigned different levels of relevance (e.g. highly relevant, partially relevant, not relevant)

Importance:

- Provides the basis for evaluation measures
- Helps improve algorithms by identifying which results satisfy user needs

. Group Relevance Judgements

In group relevance judgements, multiple assessors (users, domain experts, or crowd workers) evaluate documents for relevance.

. Pseudo Relevance Feedback (PRF):

- It is an automatic query improvement technique where the system assumes that the top-k retrieved documents from the initial search are relevant, even if not judged by the user.

- Process:

a. User submits an initial query.

b. System retrieves a ranked list of documents.

c. Top-k documents are assumed relevant.

d. System extracts terms from these docs and reformulates the query.

e. New query retrieves more accurate results.

Q.2 Explain in detail the term NDCG Explain with suitable example

Ans. NDCG stands for Normalized Discounted Cumulative Gain

- NDCG is a widely used ranking quality measure in IR and Recommender systems.
- It tells us how well a ranking system orders documents / items based on their relevance to a user query.
- Unlike precision and recall (which only consider relevant or non-relevant), NDCG considers graded relevance (e.g., highly relevant, partially relevant, not relevant).
- Step 1: Cumulative Gain (CG)

Cumulative Gain (CG) simply adds up relevance scores of retrieved documents.

If we have a ranking of results, where each document has a relevance score rel_i, then:

$$CG_p = \sum_{i=1}^{p} rel$$

where p is number of retrieved results considered (cut-off position)

- Step 2: Discounted Cumulative Gain (DCG)

To account for position, we discount lower-ranked documents (i.e. give less credit to documents that appear later).

$$DCG_p = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{\log_2(i+1)}$$

rel_i = relevance of document at position i.

The denominator $\log_2(i+1)$ pendlizes documents that appear lower in ranking.

- Step 3: Normalized DCG (NDCG)

DCG is not bounded (it depends on query length, relevance values, etc.)

To make it comparable across queries, we

normalize it by dividing with ideal DCG (IDCG) ie the best possible ordering of documents:

$$NDCG_p = \frac{DCG_p}{IDCG_p}$$

IDCG = DCG value for ideal ranking (documents sorted by highest relevance first)

Thus, $0 \leq NDCG_p \leq 1$.

Closer to 1 = better ranking

* Example

Suppose we retrieve 5 documents for a query, with following relevance scores (on a 0-3 scale, where 3 = highly relevant)

Rank	Document	Relevance
1	d1	3
2	d2	2
3	d3	3
4	d4	0
5	d5	1

- Step A: Compute DCG at $p = 5$

$$DCG_5 = rel_1 + \frac{rel_2}{\log_2(3)} + \frac{rel_3}{\log_2(4)} + \frac{rel_4}{\log_2(5)} + \frac{rel_5}{\log_2(6)}$$

$$= 3 + \frac{2}{\log_2(3)} + \frac{3}{\log_2(4)} + \frac{0}{\log_2(5)} + \frac{1}{\log_2(6)}$$

$$= 3 + \frac{2}{1.585} + \frac{3}{2} + 0 + \frac{1}{2.585}$$

$$= 3 + 1.262 + 1.5 + 0 + 0.387$$

$$= 6.149$$

- Step B: Compute IDCG at $p = 5$

Ideal order is: [3, 3, 2, 1, 0]

$$IDCG = 3 + \frac{3}{\log_2(3)} + \frac{2}{\log_2(4)} + \frac{1}{\log_2(5)} + \frac{0}{\log_2(6)}$$

$$= 3 + \frac{3}{1.585} + \frac{2}{2} + \frac{1}{2.322} + 0$$

$$= 3 + 1.893 + 1 + 0.431$$

$$= 6.324$$

- Step C: compute NDCG

$$NDCG = \frac{DCG_5}{IDCG_5}$$

$$= \frac{6.149}{6.324}$$

$$\approx 0.972$$

MRR Mean Reciprocal Rank

- It is statistical parameter
- used to evaluate any process that generates list of results for a user query
- This list should be arranged by probability of success or relevance to the user query.
- It is inversion in multiplicative form related to first answer of the query.
- Equation of MRR is

$$MRR = 1/|n| \left(\sum_{i=0}^{n} 1/\text{rank}(i) \right)$$

where

n is no. of user queries being processed

rank(i) is position of most relevant document in ith query.

b. F-score

- This is one of the most important evaluation parameter
- It is based on both precision and recall.
- F1 score = $\frac{(\text{Precision} * \text{recall}) * 2}{(\text{Precision} + \text{Recall})}$

c. F-measure

- A measure that combines recall and precision. The idea is to allow the user to specify whether he/she is more interested in recall or in precision.
- It is defined as follows

$$E(j) = 1 - \frac{1+b^2}{\frac{b^2}{r(j)} + \frac{1}{P(j)}}$$

$$= 1 - \frac{(1-b^2) \cdot P(j) \cdot r(j)}{b^2 P(j) + r(j)}$$

where

r(j) is the recall at the jth position in the ranking

P(j) is the precision at the jth position in the ranking.

It answers: Of all the negative cases, how many did the model correctly classify as negative? Important in medical testing (e.g., avoiding false alarms).

Q.15 Define and explain Interface support for search process related to visualization in information system.

Ans. In IR systems, the effectiveness of search is not only determined by ranking algorithms but also by the interface support provided to users during research process.

1. Role of Interface in Search Process

- Main goals of interface support are:-

- To assist query formulation (help users express information needs clearly)
- To visualize search results (show relevance, ranking)
- To support exploration and refinement
- To improve user satisfaction

2. Interface support Techniques

a) Query specification and Reformulation

- Interfaces provide search boxes, filters, and advanced search operations.

b) Result visualization

- Instead of just showing ranked lists, results can be shown as:
- Scatter plots or maps
- Clustered groups
- Graph based visualization

16. Calculate the precision and recall scores for the search, using the given data. A database conta
160 relevant records on a particular topic and
search was conducted on that topic and 120
rds were retrieved out of 120 records

2. Interface support techniques

a) Query specification and Reformulation

- Interfaces provide search boxes, filters, and advanced search operations

b) Result visualization

- Instead of just showing ranked lists, results can be shown as
- Scatter plots or maps
- Clustered groups
- Graph based visualization

Q.16 Calculate the precision and recall scores for the search, using the given data. A database contains 160 relevant records on a particular topic and a search was conducted on that topic and 120 records were retrieved out of 120 records retrieved, 90 were relevant.

Soln: Given Data:

- Total relevant records in db = 160
- Retrieved records = 120
- Retrieved relevant records = 90

1. Precision

$$P = \frac{\text{Relevant Retrieved}}{\text{Total Retrieved}} = \frac{90}{120}$$

$$= 0.75$$

$$P = 75\%$$

2. Recall

$$R = \frac{\text{Relevant Retrieved}}{\text{Total Retrieved in DB}} = \frac{90}{160}$$

$= 0.5625$

$= 56.25\%$

Assignment No 3

Unit III Evaluation and Visualization of Information Retrieval System

Q1 Calculate precision and recall for following example: The set of relevant documents for query q is $Rq = \{d3, d7, d8, d11, d14, d19, d23, d25\}$ where q is an information request. Using a new information retrieval algorithm, documents are retrieved as follows, The retrieved answer set is $S = \{d1, d2, d3, d7, d9, d10, d14, d20, d23, d24, d25\}$

Soln: Given:

- Relevant documents (groudnd truth):
  $Rq = d3, d7, d8, d11, d14, d19, d23, d25$
  So, $|Rq| = 8$
- Retrieved documents (system output):
  $S = d1, d2, d3, d7, d9, d10, d14, d20, d23, d24, d25$
  So, $|S| = 11$

- Step 1: Find relevant documents that were retrieved

  Intersection $Rq \cap S$

  $Rq \cap S = \{d3, d7, d14, d23, d25\}$

  True positives (TP) = 5

- Step 2: Calculate Precision

  Precision = fraction of retrieved documents that are relevant

  $$\text{Precision} = \frac{|Rq \cap S|}{|S|} = \frac{5}{11}$$

  Precision ≈ 0.4545 (45.45%)

- Step 3: Calculate Recall

  Recall = fraction of relevant documents that were retrieved

  $$\text{Recall} = \frac{|Rq \cap S|}{|Rq|} = \frac{5}{8}$$

  Recall = 0.625 (62.5%)

Information Storage and Retrieval

Unit 3 Evaluation and Visualization of Information Retrieval Systems

Nov Dec 2022

Q2. b) Define Precision and Recall. Give example of each and justify its use in evaluating IR system.

Ans Precision and Recall are two fundamental evaluation metrics used to measure the effectiveness of an Information Retrieval (IR) System, such as a Search Engine.

1 Precision

Definition - Precision is the ratio of the number of relevant documents retrieved to the total number of documents retrieved

$$\text{Precision} = \frac{\text{Relevant documents retrieved}}{\text{Total documents retrieved}}$$

e.g Suppose a search system returns 10 documents for a query, and out of those, 7 are relevant.

Precision = 7/10 = 0.7 (70%)

Precision indicates how accurate the system's results are. High precision means ~~low~~ fewer irrelevant results, which is useful when the user wants highly specific information, such as medical or legal research

2 Recall

Definition - Recall is the ratio of the number of relevant documents retrieved to the total number of relevant documents available in the system

Recall = $\frac{\text{Relevant documents retrieved}}{\text{Total relevant documents in collection}}$

e.g. A system has 50 relevant documents in the entire database for a query, and retrieves 7 of them

Recall = 7/50 = 0.14 = 14%

Recall measures how complete the search results are. High recall is important when missing relevant information is costly, such as in legal investigation, academic research, or medical diagnosis.

Both metrics are ~~are~~ crucial for a balanced evaluation.

May-Jun 2023

Q1 a) Why are the performance evaluation measures needed in IR system? Explain trade-off between Recall and Precision

Ans Performance evaluation measures are necessary to determine how effectively an IR system fulfills user needs. IR systems - such as search engines, digital libraries, or question-answering systems - must retrieve relevant information from a large collection. Evaluation metrics help in:

1. Assessing system quality - To measure how well the system retrieves relevant documents
2. Comparing different IR Models - To decide which algorithm or ranking method performs better.
3. System Improvement - Provides feedback for refining indexing, ranking, and query processing.
4. Understanding user satisfaction - Good evaluation metrics indicate whether users can find what they need easily.

5. Balancing efficiency and effectiveness - Helps tune retrieval methods for optimum performance.

Without performance measures, developers would have no scientific way to judge if changes in the system improve or degrade results.

Trade-off Between Recall and Precision

Precision and Recall are often in conflict - improving one may worsen the other. This is called the trade-off

- Precision focuses on retrieving only relevant documents.
- Recall focuses on retrieving all possible relevant documents.

When we increase recall, we retrieve more documents overall (including some irrelevant ones), so precision typically decreases.

When we increase precision, we retrieve fewer documents (eliminating irrelevant ones), so we may miss some relevant documents, reducing recall.

e.g. Suppose a query has 50 relevant documents in the collection. System retrieves:

- 10 documents, 9 relevant → Precision = 0.9, Recall = 0.18
- 40 documents, 30 relevant → Precision = 0.75, Recall = 0.60

- Precision and Recall measure different aspects of IR performance.
- A balance is required, depending on application:
  - High precision for legal, scientific, medical searches
  - High recall for research, intelligence, crime investigation
- IR systems typically use combined measures like F-score to balance the trade-off.

May - Jun 2024

Q1 a) Define Query. What are the various techniques used to specify query in information visualization.

Ans. Query - In an Information Retrieval (IR) System, a query is a formal request that a user submits to the system in order to retrieve information. It represents the information need of the user expressed through keywords, natural language sentences, or structured expressions.

e.g. If a user wants information about machine learning applications, their query might be:
machine learning applications
The IR system processes this query to retrieve relevant documents from the document collection.

* IR System Visualization

- ~~Start~~ Starting Points.

Starting points refer to the initial interface or entry point that users interact with when searching for information. They guide the user into the document space before exploration begins.

Examples of starting points:

- Search box for entering queries
- Browsing categories or menus (directory-style navigation)
- Suggested topics, popular searches, or recommended items
- Visual map of document clusters

Purpose:

- Provides users with a clear beginning for exploration
- Helps users who do not know the exact query
- Improves usability and search effectiveness

Document context

Document context refers to visual information that provides background or surrounding details about a document to help the user judge its relevance before fully opening it.

Examples:

- Snippets or summaries from the document
- Keywords or highlighted matched terms
- Citation links and related document recommendations
- Document similarity maps or cluster visualizations
- Metadata: author, date, title, source, category.

Purpose:

- Helps users quickly understand the content and decide relevance
- Reduces time spent opening irrelevant documents
- Improves retrieval accuracy and user satisfaction.