Assignment 01: Introduction

Performance summary

✓ Assessed

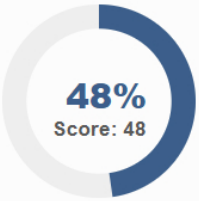
Success status	Score	Attempts
Undefined	48 of 100 points	2

Results

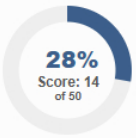
Course	**Web Retrieval SoSe 2025** ID: 4853531344 / 109642108580076
Test	**Assignment 01** ID: 4548624837

This are your test results

Duration	1h 3m 6s 4/23/2025, 3:49 PM - 4/23/2025, 4:52 PM
Answered	11 of 11 questions (100%)
Your score	48 of 100 points (48%)

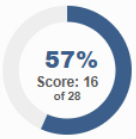


1. Knowledge Tasks (50 Points) 4



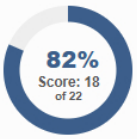
go to section

2. Practical Tasks (28 Points) 3



go to section

3. Programming Tasks (22 points) 4



go to section

1. Knowledge Tasks (50 Points) 14 of 50 points (28%)

True/false (14 Points)

Status	Answered	
Your score	14 / 14	100%

Response

Which of the following statements are correct?

Please note:

Maximum Overall Score -> 14 points

Minimum Overall Score -> 0 points

Incorrect Answer -> -2 points

Unanswered -> 0 points

Unanswered	Right	Wrong	
☐	☐	☑	In browsing, the user is more interested in look up search rather than exploratory search.
☐	☑	☐	Modern information systems use human-like languages rather than sophisticated terminologies and languages for retrieval.
☐	☑	☐	Document surrogates are used more commonly than complete documents to display the answers to a user query.

☐	☑	☐	Filtering is a retrieval task in which the information need of the user is relatively static while new documents constantly enter the system.
☐	☑	☐	Hypertext models show document relationships as edges of a generic graph in which the documents are the nodes.
☐	☑	☐	Information Retrieval is focused more on unstructured type data than structured type.
☐	☑	☐	SuperBook is a type of retrieval system which represents the structure of a large document besides the document surrogate in the answer set.

▸ Solution

IR Process (15 Points)

Status	Answered	
Your score	0 / 15	0%

Response

What are the assumptions taken into consideration for Information Retrieval? Explain how it is different from Data Retrieval. Provide at least one example to support your reasoning.

Information Retrieval makes the following assumptions: (1)users are searching for the relevant information, not the exact data, (2)content is usually unstructured or semi-structured, like web pages or articles, and (3)Relevance is subjective and could be different from person to person.

On the other side, Data Retrieval focuse on the structured data which are stored in the databases. It assumes: (1) queries are specific and return exact results, (2) data is organized in predefined formats (tables, fields), and (3) matching is binary—data either matches the query or it doesn't.

Suppose a user queries: "Benefits of drinking green tea"

In Information Retrieval: - The system searches through a collection of health-related articles or blogs. and it returns documents that discuss the health benefits of green tea, even if the exact phrase is not present. where some documents might talk about antioxidants, metabolism, or fat-burning effects – all related to the query.

In Data Retrieval: -For instance, executing a SQL query such as SELECT * FROM health_facts WHERE topic = 'benefits of green tea' will retrieve rows that exactly match that specific phrase. If no such row exists, it returns nothing, even if similar data is available under a different label (like 'green tea advantages').

195 words (max. 200)

▸ Solution

ML in IR (10 Points)

Status	Answered	
Your score	0 / 10	0%

Response

How could Machine Learning be used in Information Retrieval? Provide at least one example to support your reasoning

Machine Learning plays a powerful role in improving Information Retrieval systems by making them intelligent, faster, and more user-specific. Earlier IR systems used simple matching of keywords to retrieve the relevant documents. However, ML enables such systems to understand the meaning behind words, adapt to users' preferences, and provide more specific results.

One of the main uses of ML in IR is ranking the search results. Instead of simply counting how often a keyword appears in a document, ML models can learn from users' behavior, such click patterns, time spent on the particular pages, and past search history–to predict which results are most relevant to the user. This is called learning to rank.

ML is also used in query understanding. Users often make vague or incomplete queries. Machine learning helps the system interpret user intent, even when the exact words aren't used. This makes the search experience more natural and effective.

Example: In the university, the library's digital system could use ML to recommend research papers to students based on their past searches and downloads. If a student frequently searches for "Web Retrieval," the system can automatically suggest related papers, even if they don't contain the exact term.

Overall, machine learning makes information retrieval systems more intelligent by allowing them to learn from data and improve over time.

216 words (max. 250)

▸ Solution

IR as a Research Area (11 Points)

Status	Answered	
Your score	0 / 11	0%

Response

Discuss the merits and limitations of *system-centred* and *user-centred* approaches in Information Retrieval (IR) research. Reflect on the statement by Oard et al. (2008) that:

'IR is ultimately a human activity'

and evaluate the potential benefits of integrating the strengths of both humans and machines in the IR process. Provide examples of how such a synergy could enhance the capabilities of current IR systems and the user experience.

```
In Information Retrieval research, both system-centred and user-centred approaches have their own strengths
and weaknesses. In Information Retrieval research, system-centred approaches focus on improving algorithms,
indexing methods, and performance metrics like precision and recall. Their strength lies in efficiency,
scalability, and automation. However, they often overlook user context, preferences, and satisfaction.

User-oriented approaches often provide search results that are more aligned with individual preferences and
seem more meaningful to the user. However, they might struggle with maintaining consistency and speed when
dealing with very large datasets.

As Oard et al. (2008) pointed out, “IR is ultimately a human activity.” This refers to the reality that, no
matter how advanced the technology becomes, the primary goal of Information Retrieval is still to meet the
needs and expectations of human users.

Integrating both approaches can lead to powerful IR systems.

A great example of this is Amazon, where Amazon's recommendation system combines customers' past purchases
and browsing history with algorithm-generated suggestions to recommend products that match their
preferences. This kind of collaboration between human understanding and machine power creates smarter, more
responsive IR systems and significantly enhances the user experience.
```

191 words (max. 200)

▸ Solution

‹ go back to overview

2. Practical Tasks (28 Points) 16 of 28 points (57%)

Query and Processing (4 Points)

Status	Answered		
Your score	4 / 4		100%

Response

The process that transforms a document's “original” representation into the representation that is considered for searching is called “**Pre-processing**”. And the process that parses the human search query into a machine-readable representation is called “**Query Processing**”.

You are given two examples. Which do you think is an example of a “Query Processing” and which one belongs to “Pre-processing”?

Each correct answer scores 2 points; each incorrect answer results in a deduction of 1 point.

	Query Processing	Pre-Processing
Query: “Information Retrieval -Web” **Processed: “Information” AND “retrieval” AND NOT “web”**	☑	☐
Original: “ Web Information Retrieval Searching the web… ” **Processed: [“web”, “information”, “retrieval”, “searching”, “web”, …]**	☐	☑

▸ Solution

IR Architecture (12 Points)

Status	Answered		
Your score	12 / 12		100%

Response

Using an example of a query related to education, such as "study tips for biology exams," put in order the steps an IR system would take to process this query and present relevant results.

Results Presentation: A list of relevant documents is displayed to the user, often with helpful excerpts.

Query Parsing: It breaks down the query into keywords like "study," "tips," "biology," and "exams."

Relevance Ranking: The system sorts the found documents by how well they match the query.

Step 1

Query Reception: The IR system receives the input query "study tips for biology exams."

Step 2

Query Parsing: It breaks down the query into keywords

like "study," "tips," "biology," and "exams."

Query Expansion: The system adds related terms to the query to broaden the search.

Query Reception: The IR system receives the input query "study tips for biology exams."

Search Execution: It searches its database for documents containing the query terms.

Step 3

Query Expansion: The system adds related terms to the query to broaden the search.

Step 4

Search Execution: It searches its database for documents containing the query terms.

Step 5

Relevance Ranking: The system sorts the found documents by how well they match the query.

Step 6

Results Presentation: A list of relevant documents is displayed to the user, often with helpful excerpts.

▸ Solution

IR Design (12 Points)

Status	Answered	
Your score	0 / 12	0%

Response

Imagine you are designing an IR system for a university's academic resource center. Describe three aspects that you would take into consideration for addressing the specific information needs of students looking for research articles in their field of study.

Tip: Some examples of such aspects can be the ranking algorithm, the creation of user profile, the presentation of search results, etc. If you cannot think of other aspects, you can explain these examples.

```
Designing an IR System for a University's Academic Resource Center

When designing an Information Retrieval system for a university's academic resource center, it is essential
to consider the specific needs of students searching for research articles. Here are three important
aspects:

Ranking Algorithm
.An effective ranking mechanism helps display the most useful research paper first. Such a mechanism is
capable of parsing various parameters, including keyword relevance, publication date, citations, and
authors' credibility. For example, when a computer science student is searching for literature on machine
learning applications in healthcare, the system ought to rank the most current and most relevant papers
that specifically connect the two subjects.

User Profile Creation Building user profiles based on students' academic history, previous questions, and
bookmarked sources allows the system to provide personalized results. For instance, a data science student
and a biology student searching for "genomics" can receive different sets of results tailored to each's
field.

Presentation of Search Results

Results should be displayed clearly, with filters such as publication date, field, author, or document
category (e.g., journal articles, conference proceedings, or theses). Abstract previews, keywords, and
direct download or citation options can increase navigational efficiency and convenience. These features
together help students at universities like University of Koblenz find relevant academic resources more
effectively.
```

206 words (max. 250)

▸ Solution

go back to overview

3. Programming Tasks (22 points) 18 of 22 points (82%)

You may have heard of the Titanic accident. Here, we use Python programming to explore the data for gaining further insights. You can access the dataset here http://web.stanford.edu/class/cs102/datasets/Titanic.csv

True/false (12 Points)

Status	Answered	
Your score	12 / 12	100%

Response

Which of the following statements are correct?

Please note:

Incorrect Answer -> -2 points

Unanswered -> 0 points

Unanswered	Right	Wrong	
☐	☐	☑	The dataframe does not contain any null values (2 points)
☐	☐	☑	Only two columns of dataframe contain null values (3 points)
☐	☐	☑	There are more survived male passengers than female (3 points)
☐	☐	☑	More male passengers survived than died (2 points)
☐	☑	☐	The chances of survival are lower if the passenger had bought the cheapest ticket, i.e. travelling in a lower class (3 being the least and 1 being the best) (2 points)

▸ Solution

☑ Plotting (2 Points)

Status	Answered	
Your score	0 / 2	0%

Response

Which of the following line of code is correct to achieve the following:

Let's assume that we want to plot analysis by gender indicating the men, women in a bar chart.

True	False	
☑	☐	dataframe.value_counts('gender').plot()
☐	☑	dataframe['gender'].value_counts().plot().bar()
☐	☑	dataframe['gender'].value_counts().plot.bar()
☐	☑	dataframe.value_counts()

▾ Solution

Which of the following line of code is correct to achieve the following:

Let's assume that we want to plot analysis by gender indicating the men, women in a bar chart.

True	False	
☐	☑	dataframe.value_counts('gender').plot()
☐	☑	dataframe['gender'].value_counts().plot().bar()
☑	☐	dataframe['gender'].value_counts().plot.bar()
☐	☑	dataframe.value_counts()

••• SQL like queries (2 Points)

Status	Answered	
Your score	0 / 2	0%

Response

Write the (SQL like) query for dataframe to retrieval records of males ('M') survived

Note: the dataframe's name is "dataframe".

```
dataframe[(dataframe['gend
```

▾ Solution

Write the (SQL like) query for dataframe to retrieval records of males ('M') survived

Note: the dataframe's name is "dataframe".

```
dataframe.query("gender ==
```

dataframe.query("gender == 'm' and survived == 'Yes'", dataframe.query("gender == 'M' and survived == 'yes'", dataframe.query("gender == 'm' and survived == 'yes'"

Conditional probability (6 Points)

Status	Answered	
Your score	6 / 6	100%

Response

What are the conditional probability of being survived given the gender and passenger class?

- 0.97
- 0.37
- 0.16
- 0.5
- 0.92
- 0.14

P (Survived | female, class : 1)
0.97

P (Survived | female, class : 2)
0.92

P (Survived | female, class : 3)
0.5

P(Survived | male, class:2)
0.16

P(Survived | male, class : 3)
0.14

P(Survived | male, class:1)
0.37

▸ Solution

go back to overview

Test execution

Information

Availability: Expired at 4/24/2025, 1:59 PM

Max. attempts: Unlimited

Results of this test are visible to administrators and tutors of this course.

Start test

▸ Change log

Go to top