

Q1

1/1

When we upload pics/clips/songs... to social media, **what specific mechanism** to we use, to help others find our content when they search?  
Why do we need this specific mechanism?

---

Q2

1/1

What algorithms causes/leads to/results in/is implicated in... **'filter bubbles'**?  
How does it lead to this?

---

Q3

1/1

**What TWO other items** can a search engine serve us (eg. via 'snippets'), in addition to what gets served already? Name each item, and briefly state why it would be of use to us.

---

Q4

1/1

As you know, genAI (generative AI) is so-called because it can generate content (text, images, video, audio, more).  
**HOW will this** adversely affect search in the (near, even) future? Explain carefully (don't write a vague answer!).

---

Q5

1/1

ChatGPT (for example) is said to "hallucinate" sometimes (or a lot of times, depending on the type of questions) - an unfortunate term (because only minds can hallucinate!) used by companies who serve this kind of AI products (eg.

<https://fortune.com/2023/04/17/google-ceo-sundar-pichai-artificial-intelligence-bard-hallucinations-unsolved/>). This means that the bot provides an incorrect answer (which we can verify using our own knowledge or experience or by doing a good old search!).

**WHAT mechanism (in the algorithm) causes this to happen?** Please be specific.

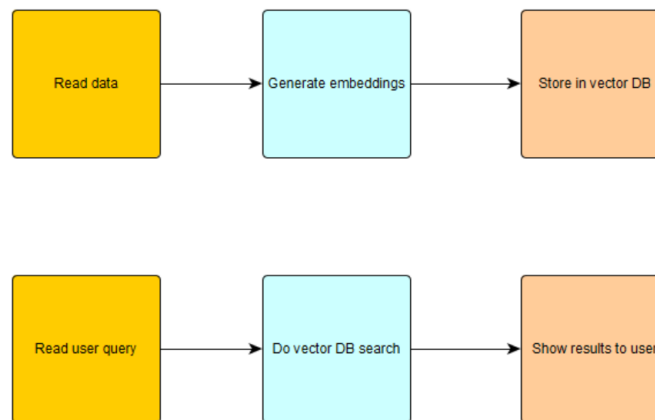
---

Q6

1/1

We typically write code (eg for your HWs #2 to #5) to make use of IR algorithms.

An alternative way is to use 'nodes' (a node is a box-like representation that encapsulates a specific task by executing that task's code) and WIRE them up visually, like so (see for example, <https://www.google.com/search?q=rapidminer+dataflow&tbm=isch>):



**WHAT would be TWO** specific (and different from each other) advantages of switching to this way of working (using nodes, as opposed to coding)?

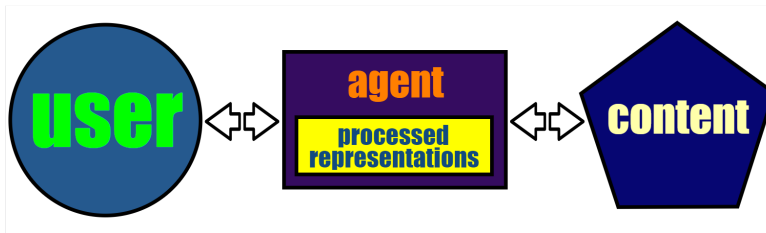
---

Q7

1/1

Consider the diagram below (the bottom part is simply a slightly zoomed-in portion of the top):





Now that the course is over (after you get through this exam, lol), how does it summarize (encapsulate) the course? **Pick four specific and different IR tasks** we studied during the course (including during the 'Assorted topics' pair of lectures), and explain (in a line or two) each, in terms of the three pieces of our diagram.

---

Q9

4 / 4

1+1=2 points: **How** do recommendation engines (REs) work?

1+1=2 points: **What** are two different uses for them when we search?

---

Q10

1 / 1

**What is 'vector similarity search'?** Rather than Googling or ChatGPT-ing, answer, based on what we covered.

For **what two different** IR tasks are vector DBs useful? Name, and explain briefly why we couldn't do them without vector DBs.

---

Q11

1 / 1

**Summarize ANY TWO** of your HWs #2 through #5 - **WHAT was the algorithm** underlying, **WHAT task** did it help accomplish?

---

Q12

1 / 1

**What** does 'OPL' stand for, in OPL stack?

**What is** its main use? Again, stick to what we covered, rather than searching online!

---

Q13

1 / 1

**How do** 'RDF triples' make search better? Explain in a few lines.

---

Q14

4 / 4

**Name four algorithms** we looked at, for IR tasks, that rely on iteration or recursion. For each, **explain briefly, how** the iteration or recursion helps (ie. what changes during each run).

---

Q15

4 / 4

**Name, and very briefly discuss** 4 ML-based algorithms we looked (towards the end of the course!), for IR tasks.

---

Q16

4 / 4

Consider the following summary of ML/DS algorithms ([https://bytes.usc.edu/cs572/s23-sear-chhh/extras/docs/KNIME\\_ML\\_cheatsheet.pdf](https://bytes.usc.edu/cs572/s23-sear-chhh/extras/docs/KNIME_ML_cheatsheet.pdf)):

Cheat Sheet: Machine Learning with KNIME Analytics Platform

**SUPERVISED LEARNING**

**CLASSIFICATION**

**NUMERIC PREDICTION & CLASSIFICATION**

**NUMERIC PREDICTION**

**TIME SERIES ANALYSIS**

**UNSUPERVISED LEARNING**

**CLUSTERING**

**END-TO-END LEARNING**

**BAGGING**

**BOOSTING**

**DEPLOYMENT**

**EVALUATION**

**RECOMMENDATION ENGINES**

Of the various algorithms listed above, **pick FOUR** that are useful in IR [we studied them], and explain briefly how each works: **what** does the algorithm do, **what IR task** does it help with.

Q17

1/1

TikTok's recommendation engine uses a specific data structure, to optimize how it works. **What** is the name of the data structure? In your own words, **how** does it work (you can explain a high level, no need for specifics)?

Q18

1/1

Given two vectors (eg like shown below), **what two** 'similarity measures' can we calculate ?  
**What** do vectors have to do, with IR in the first place?!