Web Information Retrieval
Q&As: Preparation for the exams
SOSE2023

Frank Hopfgartner, Stefania Zourlidou
Institute for Web Science and Technologies

Topics

- Answer questions posted in padlet

- Answer questions regarding the course, the assignments and the exams

- Discuss a demo exam paper

can you please explain the calculation of rocchio feedback programming question (similar to the assignment) using an example?

💬 1

👤 **zourlidou** 2h   ⋮
yes, I will do it today in the lecture

👤 Add comment

---

could you describe the exam pattern? what weightage will be given to each topic that we have covered during weekly assignments?

💬 1

👤 **zourlidou** 2h   ⋮
I will show a demo test today, so it will become very clear the format of the exam paper. About the weights of the questions: I cannot say, as the exam questions are not yet deided. They will be certainly many topic categories to cover a wide range of topics discussed in the lectures, so the weights will be shaped accordingly.

👤 Add comment

Hi, are we expecting any programming questions in the paper? Also will there be any demo test? Will the question patterns be like the assignments?

💬 3

👤 **zourlidou** 2h   ⋮
Yes, I will show a demo test in the lecture session today. Programming questions may be given, but they will be a matter of writing a few lines codes. And yes, the questions pattern or format will be very similar with that of assignments.

👤 **Anonymous** 1h   ⋮
Could you give a general example of how a Programming question could look like in the exam? Or would it be very similar to the questions in the Assignments?
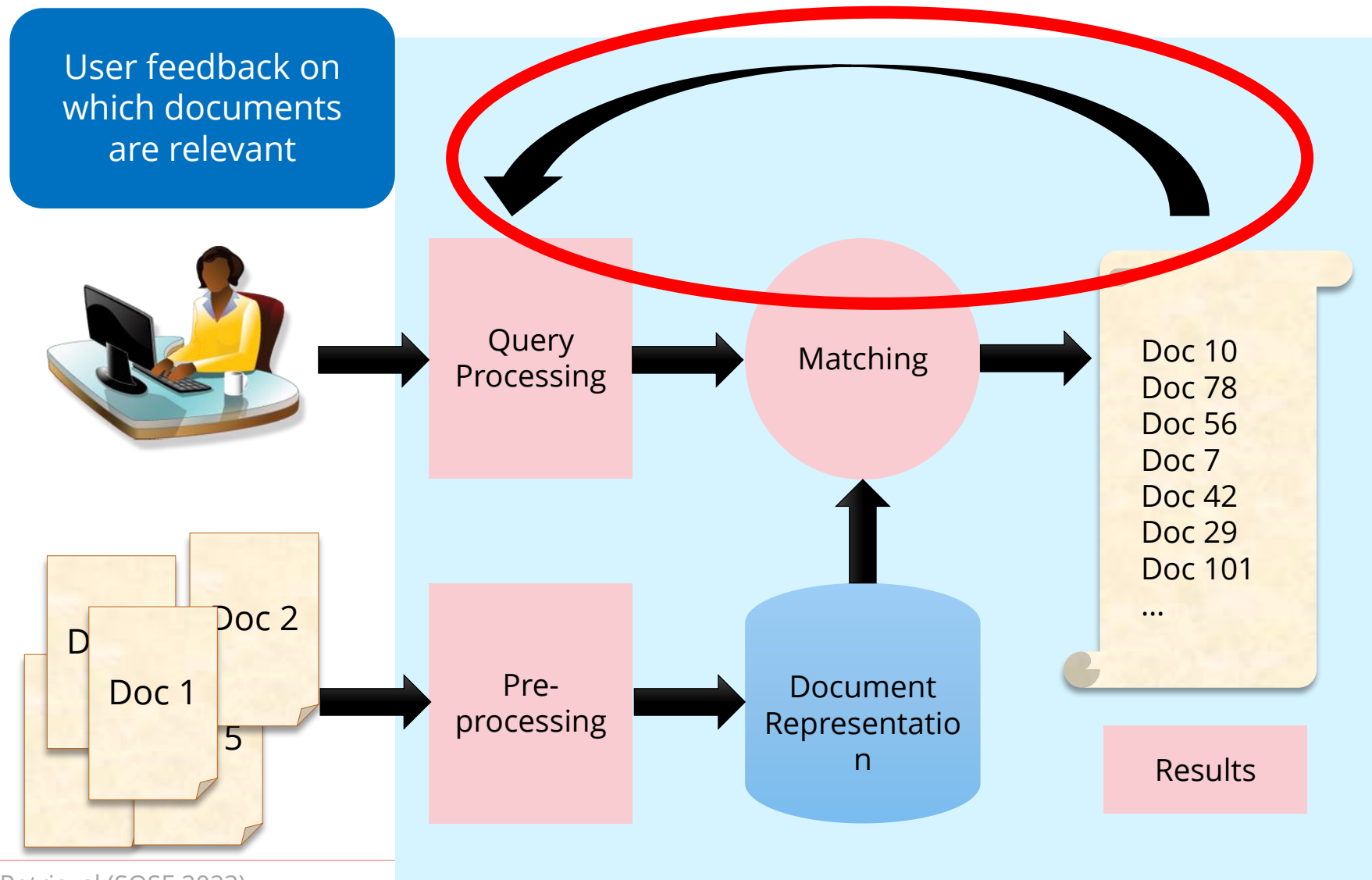
👤 **zourlidou** 1m   ⋮
It can be write a function in python, or code will be given where there are some gaps that mush be filled in from a given list (I will show a demo today at the lecture at 14:00)

👤 Add comment

3. Relevance Feedback (Rocchio)

Relevance Feedback for VSM

User feedback on which documents are relevant

Query Processing

Matching

Pre-processing

Document Representation

Doc 10
Doc 78
Doc 56
Doc 7
Doc 42
Doc 29
Doc 101
...

Results

Doc 2
Doc 1

Incorporate Feedback

- User provides for some documents if they are relevant or irrelvant:

  o D^p : set of documents with positive feedback (relevant)

  o D^n : set of documents with negative feedback (irrelevant)

- Adjust query vector

$$\vec{q}_{FB} = \alpha \cdot \vec{q} + \beta \frac{1}{\left|D^p\right|} \sum_{d_i \in D^p} \vec{d}_i - \gamma \frac{1}{\left|D^n\right|} \sum_{d_i \in D^n} \vec{d}_i$$

- Parameters: typically $\alpha > \beta > \gamma$

- Adjust negative term weights to 0

Example

- Query:
  - paris hilton

- Documents

	paris	hilton	hotel	france	eiffel	blonde	heiress	actress
d_1	3	1	1	2	1			
d_2	1	3	4	1	3			
d_3	2	1		1				
d_4	3			2			1	
d_5		1	3					
d_6	3	3				2		
d_7	2	2					2	
d_8	2	1	1			1	1	
d_9	3	2				1		4
d_{10}	3	2		1			2	3

Example

- Initial result list based on VSM

- User provides relevance feedback

  - Positive 🙂

  - Negative 🙁

- Parameter:

  - $\alpha = 1$

  - $\beta = 0.75$

  - $\gamma = 0.15$

Rank	Doc	ρ	
1	d_3	0.395	🙂
2	d_6	0.183	
3	d_7	0.161	🙁
4	d_4	0.132	
5	d_1	0.128	🙂
6	d_8	0.125	🙁
7	d_{10}	0.071	
8	d_9	0.057	
9	d_2	0.049	
10	d_5	0.027	

Example

- Adjust query

$$\vec{q}_{FB} = \alpha \cdot \vec{q} + \beta \frac{1}{|D^p|} \sum_{d_i \in D^p} \vec{d}_i - \gamma \frac{1}{|D^n|} \sum_{d_i \in D^n} \vec{d}_i$$

α = 1
β = 0.75
γ = 0.15

$$\vec{q}_{FB} = 1 \cdot \begin{pmatrix} 0.046 \\ 0.046 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + 0.75 \cdot \frac{1}{2} \cdot \left[\begin{pmatrix} 0.137 \\ 0.046 \\ 0.398 \\ 0.602 \\ 0.699 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.092 \\ 0.092 \\ 0 \\ 0.301 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right]$$

$$- 0.15 \cdot \frac{1}{2} \cdot \left[\begin{pmatrix} 0.092 \\ 0.092 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.796 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.092 \\ 0.046 \\ 0 \\ 0 \\ 0 \\ 0.523 \\ 0.398 \\ 0 \end{pmatrix} \right] \mapsto \begin{pmatrix} 0.118 \\ 0.087 \\ 0.119 \\ 0.339 \\ 0.262 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Rank	Doc	ρ
1	d_1	0.957
2	d_3	0.787
3	d_2	0.691
4	d_4	0.640
5	d_5	0.262
6	d_8	0.172
7	d_{10}	0.119
8	d_6	0.056
9	d_7	0.050
10	d_9	0.018

Other questions?

Demo exam paper

References

[1] https://olat.vcrp.de/auth/RepositoryEntry/4071063853

[2] https://nlp.stanford.edu/IR-book/information-retrieval-book.html C. D. Manning, P. Raghavan and H. Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.

[3] S. Ceri, A. Bozzon, M. Brambilla, E. Fraternali, S. Quarteroni, Web Information Retrieval, Springer-Verlag Berlin Heidelberg, 2013.

- Chapter 13, Multimedia Search, available in OLAT.