Fairness and Transparency in IR

[DAT640] Information Retrieval and Text Mining

Nolwenn Bernard
University of Stavanger

October 31, 2022

In this module

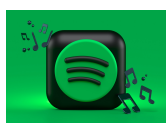
1. Sociotechnical system
2. Fairness in IR
3. Transparency in IR

Sociotechnical system

Applications of IR

- Search engines
- Product and items recommendation, *e.g.*, music, movies, books, news, etc.
- Candidate ranking, *e.g.*, hiring platform, college admission, etc.
- Etc.



Position of the people

- Where do the people intervene in IR applications?
  - Algorithms and systems design, implementation, and evaluation (*e.g.*, developers, annotators)
  - Consumers
  - Producers

Position of the people

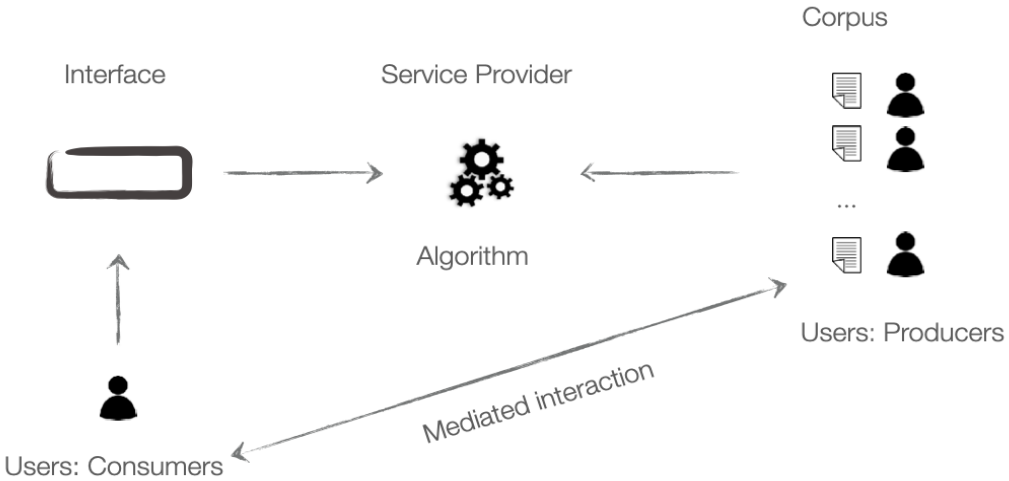


Figure: Information Retrieval system. Figure taken from Biega talk at ESSIR '22.

Sociotechnical system

- A IR system is a **sociotechnical system** due to the *omnipresence of people* in its ecosystem
  - It is used by producers and consumers
  - It learns from users behavior and feedback
  - It is designed and implemented by people
  - It can have an impact on the society, *e.g.*, consumption habits, political outcomes
  - Etc.

Negative impacts

- **Distributional harms**: unfair distribution of a resource to an individual or group of consumers, *e.g.*, unfair exposure of producers
- **Representation harms**: underrepresentation of a category of items from the corpus, *e.g.*, independent movies, women
- **Privacy loss**: collection of consumers personal data, disclosure of producers information

Fairness in IR

Question

What does it mean to be fair?

Common definition

Fairness: "the quality of treating people equally or in a way that is reasonable."[1]

- *Broad* definition which can be *interpreted differently* depending on the people
- It raises several questions such as:
  - What is the goal? → Definition
  - Is the goal completed? → Evaluation
  - How can we complete the goal? → Method, implementation
- Taxonomies have been proposed to formalize fairness

[1] `https://www.oxfordlearnersdictionaries.com`

Fairness taxonomy

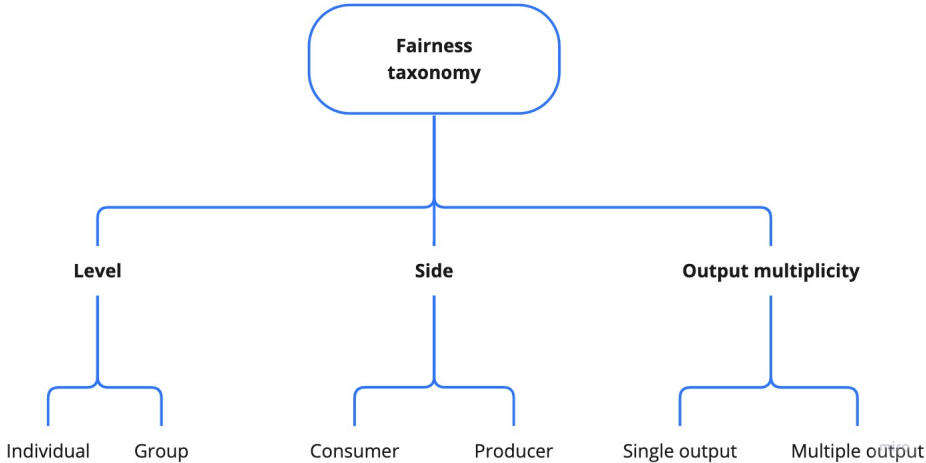


Figure: Fairness definition taxonomy (adapted from Pitoura et al., 2021)

Fairness dimensions

Level

- **Individual**: ensure consistent treatment of similar entities
- **Group**: ensure consistent treatment of entities belonging to a group

Example

Search Query	Work Experience	Education Experience	Candidate	Xing Ranking
Brand Strategist	146	57	male	1
Brand Strategist	327	0	female	2
Brand Strategist	502	74	male	3
Brand Strategist	444	56	female	4
Brand Strategist	139	25	male	5
Brand Strategist	110	65	female	6
Brand Strategist	12	73	male	7
Brand Strategist	99	41	male	8
Brand Strategist	42	51	female	9
Brand Strategist	220	102	female	10
		...		
Brand Strategist	3	107	female	20
Brand Strategist	123	56	female	30
Brand Strategist	3	3	male	40

TABLE I: Top k results on www.xing.com (Jan 2017) for an employer's job search query "Brand Strategist".

Figure: Example taken from Lahoti et al., 2019

- Group fairness is achieve in top-10
- Unfair to individuals with similar qualifications

Fairness dimensions

Side

- **Consumers**: ensure that similar users or group of users receive similar output
  - Example: democrats should have similar results when looking for information on gun regulations
- **Producers**: ensure that similar items or group of items are ranked/recommended in a similar way
  - Example: in candidate ranking, gender should not influence the final ranking

Fairness dimensions

Output multiplicity

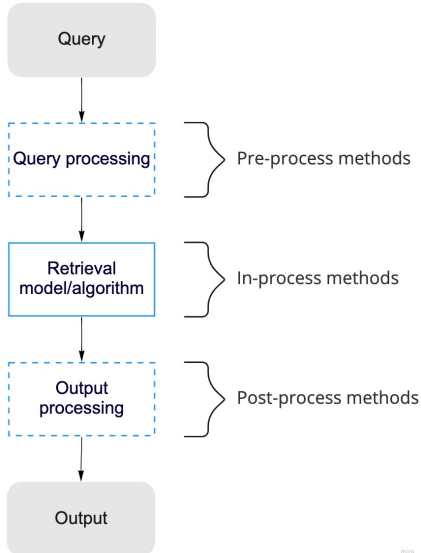
- **Single**: fairness is studied on only one output
- **Multiple**: fairness is studied on a sequence of outputs as a whole
  - Some rankings can be considered unfair in a fair system

Approach type

- Different types of solution exist to tackle the problem of fairness
  - **Model specific**: the approach does depend on the model used for information retrieval
  - **Model agnostic**: the approach does not depend on the model used for information retrieval

Approach position

- **Pre-process**: the approach is used before the retrieval process
- **In-process**: the approach modifies the retrieval process
- **Post-process**: the approach modifies the output of the retrieval process



Current trends

- Most of the literature focus on:
  - Group fairness
  - Single output
  - Item-side fairness
- Difficulty to establish state-of-art approach due to the number definitions
- A common approach is *post-process* and *model agnostic*

Exercise

E14-1 Measuring fairness in ranking

Transparency in IR

Common definition

Transparency: "the quality of something, such as a situation or an argument, that makes it easy to understand."[2]

- *Broad* definition which opens the door for many approaches
- Operationalization raises similar questions as for fairness

[2] https://www.oxfordlearnersdictionaries.com

Question

How can we make a system transparent?

Modality

Several modalities are possible to communicate with the users:

- **Article**: communication information in an article, or a blog post
  - Example: PubMed Best Match algorithm (Fiorini et al. 2018)
- **Open source**: build the system with open source resources
- **User Interface (UI)**: communicate with the user through the UI (*e.g.*, explanations, icons)

Example

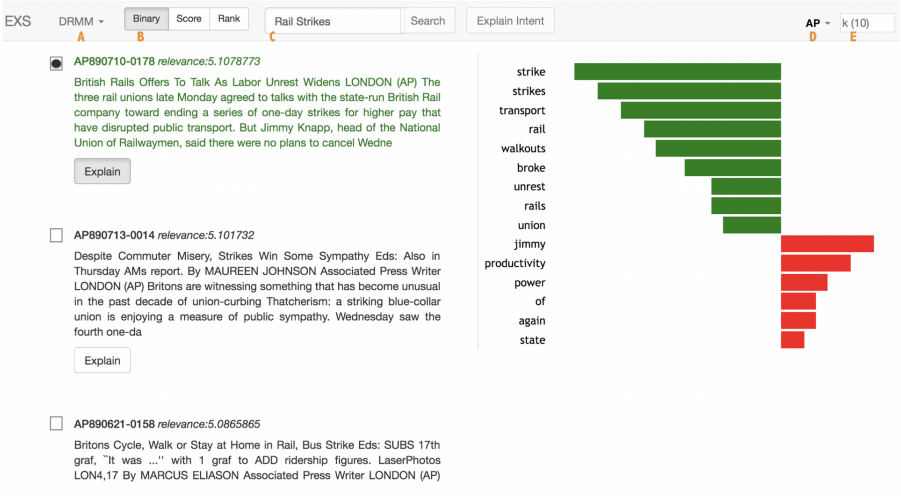


Figure: Example of the EXS user interface (Singh and Anand, 2019)

Challenges

There are many challenges related to transparency of the system, such as:

- How to be as transparent as possible without revealing business secrets to competitors?
- How to be transparent and prevent subsequent malicious behavior?
  - Example: search engine optimization, *i.e.*, website optimization to get a better ranking from the search algorithm
- Transparency of blackbox models
  - Example: explain a ranking produced by a transformer model

Transparency taxonomy

- Based on previous work and existing challenges, we identify two other dimensions for transparency:
  - Degree
  - Level

Degree of transparency

Degree

- **Full**: provide all the information necessary to make the entire system transparent
  - Example: description of the indexing and BM25 retrieval model for a search engine
- **Partial**: provide all the information necessary to make a part of system transparent
  - Example: description of the indexing, BM25 retrieval model, and transformer-based re-ranker for a search engine
  - The transformer-based re-ranker is not transparent yet

Level of transparency

Level

- **Global**: describe the inner-workings of the system
- **Local**: describe the relationship between a specific input and output
- **Causality**: describe the relationship between the inner-workings of a system (*i.e.*, the cause) and a specific output (*i.e.*, the effect)

Current trends

- A common approach is *post-process* and *model agnostic*
- Work in the literature tend to favor:
  - Local explanations
  - Communicate explanations through the UI

Exercise

E14-2 Ranking explanation

Summary

- IR systems are sociotechnical systems
  - Position of the people
  - Negative harms (incl. distributional and representational harms, privacy loss)
- Fairness in IR
  - Fairness taxonomy (incl. level, side, output multiplicity)
  - Fairness approach categorization
  - Current trends
- Transparency in IR
  - Transparency dimensions (incl. modality, degree, level)
  - Challenges
  - Current trends

Reading

- Pitoura et al. **Fairness in rankings and recommendations: an overview**. In: *The VLDB Journal (2021)*
  - `https://link.springer.com/article/10.1007/s00778-021-00697-y`
- Talk by Asia J. Biega at ESSIR '22. *Responsible Design of Information Access Systems*
  - Video: `https://www.youtube.com/watch?v=BnoiNTsUNZU`
  - Slides: `http://essir2022.org/slides/asia-biega.pdf`