



Measuring personalization in environmental search



Group 15: Alma Nilsson, Diego Bartoli
Geijo, Marta Españó López





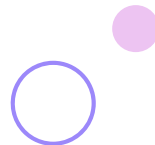
Importance of personalization

“

Does personalization performed by a search engine on user profiles, with different political views, affect the

”

results of queries related to environmental issues?



METHOD

PREPARATION

- Bing
 - **Logged browsing**
- 10 users RW & 10 LW
 - **Females, 1992, Spain**
 - **VPN, Madrid**
- 80 training queries
 - **4 categories**
 - **News, Lifestyle, Humanities, Other**
 - **20 queries each**
- 10 testing queries
 - **5 general, 5 biased**

IMPLEMENTATION

Browsing sessions:

- 1 right, 1 left
- 3 days, 2 sessions per day
- 2 random queries per category
- 1 URL from top 50% ranked results clicked per query
- Between 5 and 10 sec spent on each URL
- URLs saved for evaluation

EVALUATION

Environmental browsing sessions:

- All queries for all users
- Three sessions
 - **Before, during, after**
- 10 top results saved compared
- Two methods of comparison
 - **Jaccard similarity index**
 - **Kendall's (Tau) rank correlation coefficient**



EXPERIMENTAL SETUP - I

Always the same machine: **Dell PC** with **Windows 10**.

To ensure Madrid location: **Freedome VPN**.



To implement automated logged browsing: **selenium** (**python** library) with **chrome driver**.

Some initial technical problems, the **sessions** were **too long** causing **connection problems** \Rightarrow **3 days** of **unstructured training** and **3 days** of **structured training**.



EXPERIMENTAL SETUP - II

Session file structure

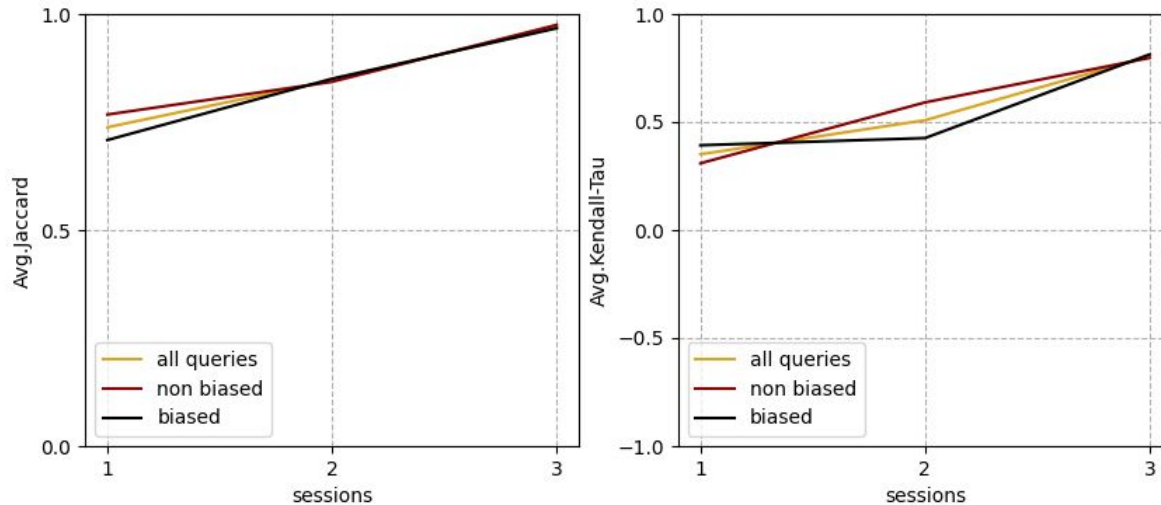
```
{" user id ": {  
  "username": "username val.",  
  "password": "password val.",  
  " session id ": {  
    " query ": [ "links vis."]  
    ....  
  }  
}  
...  
}
```

Expl. analysys training data

	Session	Wing	Missing val.	Uniq.queries	Uniq.links
1	1	left	0.0	0.9625	0.9625
2	1	right	0.0	0.9875	0.9625
3	2	left	0.2	0.8	0.8
4	2	right	0.1	0.875	0.9
5	3	left	0.0	0.975	1.0
6	3	right	0.0	1.0	1.0
7	4	left	0.0	0.9625	1.0
8	4	right	0.0	0.9875	1.0
9	5	left	0.0	0.975	1.0
10	5	right	0.0	0.9625	1.0
11	6	left	0.0	0.975	1.0
12	6	right	0.0	0.875	0.9

RESULTS

Similarity comparison for environmental (test) queries



Statistics for each session

	Jaccard	Kendall's
Session 1 env RW-LF 1	0.7378	0.3507
Session 2 env RW-LF 2	0.8467	0.5083
Session 3 env RW-LF 3	0.9709	0.8049
Session 1 RW	0.7711	0.3689
Session 2 RW	0.8955	0.6673
Session 3 RW	0.9570	0.7591
Session 1 LW	0.7214	0.3384
Session 2 LW	0.8084	0.4739
Session 3 LW	0.9814	0.82024

CONCLUSIONS

1

Profiles become more similar over time.

2

Factors that help to personalize are kept constant (e.g. name, age).

3

The main field of the query may play a more important role than the detail.

4

Political personalization may not influence environmental query search.

Future work

1

Include male and female users.

2

Click more URLs, add more variation on the time and duration of the sessions.

3

Train the profiles for a longer time.



Thank you.

