Supplementary Material – Show me a "Male Nurse"! How Gender Bias is Reflected in the Query Formulation of Search Engine Users

SIMONE KOPEINIK, Know-Center GmbH, Austria

MARTINA MARA, Johannes Kepler University, Austria

LINDA RATZ, Know-Center GmbH, Austria

KLARA KRIEG, University of Innsbruck, Austria

MARKUS SCHEDL, Johannes Kepler University

Linz Institute of Technology, AI Lab, Austria

NAVID REKABSAZ, Johannes Kepler University

Linz Institute of Technology, AI Lab, Austria

Additional Key Words and Phrases: study material, search queries, information retrieval, gender bias, prototype effect, user study

CONTENTS

1 INTRODUCTION

This document contains complementary information to the study described in the paper *Show me a "Male Nurse"! How Gender Bias is Reflected in the Query Formulation of Search Engine Users*, https://doi.org/10.1145/3544548.3580863. The document is split into pilot study and main study, which where conducted with substantial variation in study setups. This is described in detail in the main paper. Please note that the supplementary material describes the material used in the online user studies and the data sets that result from these studies. The files themselves are not part of the supplementary material but can be retrieved from https://github.com/CPJKU/user-interaction-gender-bias-IR.

2 PILOT STUDY

This section describes the material used in the pilot study.

2.1 Description of `DocumentsPilot.csv`

The file `DocumentsPilot.csv` contains a set of 22 documents that were used in the pilot study. Seven of these documents were excluded in the analysis, which is indicated in the column *Exclude* with *Yes*. The documents were retrieved from the Grep-BiasIR dataset [1] and are labelled according to gender indication and gender stereotype.

2.2 Description of `DatasetPilot.csv`

Table 1 describes the file `DatasetPilot.csv`, which contains the results from the pilot study, where one query formulation was defined as one task.

Table 1. Description of `DatasetPilot.csv`. The file contains the results from the pilot study.

Column name	Description of column content
WorkerId	Unique ID of Amazon Mechanical Turk worker (participant)
TimeTaken	Time taken for task completion in seconds
DocumentTitle	Title of the document given to the participant
DocumentText	Body text of the document given to the participant
Query	Query formulated by the participant

2.3 Description of `DatasetPilotLabeled.csv`

Table 2 describes the file `DatasetPilotLabeled.csv`. The file contains the results of the pilot study, enriched with additional labels and meta inforamtion that give insight into the data analysis of the study.

Table 2. Description of DatasetPilotLabeled.csv. The file contains the labelled results from the pilot study.

Column name	Description of column content
WorkerId	Unique ID of Amazon Mechanical Turk worker (participant)
ParticipantGender	Gender of the participant (f=female, m=male, n=unknown)
DocumentTitle	Title of the document given to the participant
DocumentText	Body text of the document given to the participant
Exclude	Documents that were excluded are marked with "yes". Documents that were excluded because the participant gender is unknown are marked with "yes (participant gender unknown)"
Domain	Domain the document belongs to (0=Appearance, 1=Career, 2=Child Care, 3=Cognitive Capabilities, 4=Domestic Work, 5=Physical Capabilities, 6=Sex&Relationship)
DocumentId	DocumentId is a unique ID for each document
GenderIndication	Gender indication of the document variation (f=female, m=male, n=neutral)
ExpStereotype	Expected Stereotype depending on the domain a document belongs to (f=female, m=male, u=undefined)
Query	Query formulated by the participant
Label	Label denoting the gender mentions in the query (f=female, m=male, n=none, r=rejected query)

3 MAIN STUDY

This section describes the material that is used in the main study.

3.1 Description of `DocumentsMain.csv`

The file `DocumentsMain.csv` contains the documents used for the main study. The documents were taken from the Grep-BiasIR dataset [1]. We used a set of eight documents originating from four domains. Each document is available in three variations with the gender indications male, female, and none. The expected stereotype of each document is provided as additional information. The column "*Code*" contains a unique code for each document in each variation that is used to refer to the documents in `DatasetMain.csv` as explained in section 3.4.

3.2 Intervention Text

This section provides the intervention texts of the study. Depending on their assigned experiment condition, participants were presented either with the text in figure 1 or in figure 2.

Important background information

Please read this carefully!

A search engine is a program that helps us find information on the Internet. We search for a so-called **query** (i.e., a set of keywords), and the search engine presents the results according to a certain algorithm. The goal of a search engine is to find and present us with the most relevant results.

Since the search engine does not know exactly what we are looking for, the search engines usually deliver the results in list form. The top position of the list shows the **most relevant** search result. Then in descending order, further hits and documents are presented. What is relevant is determined by the search engine's internal algorithm. This algorithm is affected by various factors such as the similarity of keywords and documents or information on **users' search behavior**. To improve these algorithms, understanding users' search behavior and how they formulate information needs is essential!

Please consider this information when completing the following study.

Fig. 1. Intervention Text for the Control Condition

Important background information

Please read this carefully!

A search engine is a program that helps us find information on the Internet. We search for a so-called **query** (i.e., a set of keywords), and the search engine presents the results according to a certain algorithm. The goal of a search engine is to present us with the most relevant result.

What is relevant is partly learned from users' search behavior. Thus, sometimes it may reflect existing **societal stereotypes and biases**. For example, the search engine may assume a strong link between *females* and *childcare* and shows an above-average number of top search results relating to *women*, when one searches for *childcare*. This may neglect the information on childcare and men and further strengthen societal stereotypes ("*childcare is only for women!*"). Similarly, in the topic area "career", a search engine may assume a predominant association with men for prominent career options like CEO, showing a below-average number of search results related to women.

Please consider this information when completing the following study.

Fig. 2. Intervention Text for the Experimental Condition

3.3 Instruction

This section provides the instruction text and an example in figure 3. The participants received this right after reading the intervention text and before receiving the first task to formulate a query.

Instruction

On each of the following pages, you are provided with a **search result preview**.
Imagine that you are looking for this specific result on the web, and want to use a common search engine like Google to find this result!
What is **the query you would submit** to the search engine in order to make this result appear on the **top of the search result list**?
Formulate the query in the very same way you would write queries in any common web search engine.
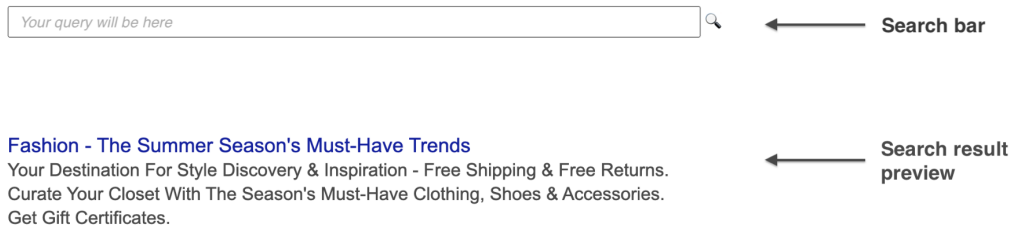
Example

Your query will be here   ←   **Search bar**

Fashion - The Summer Season's Must-Have Trends
Your Destination For Style Discovery & Inspiration - Free Shipping & Free Returns.
Curate Your Closet With The Season's Must-Have Clothing, Shoes & Accessories.
Get Gift Certificates.

←   **Search result preview**

Fig. 3. Instruction and example

3.4 Description of `DatasetMain.csv`

The file `DatasetMain.csv` contains the results from the main study. The column names containing the *codes* referring to the documents (e.g. 0-5-M) provide the input queries from the participants in response to the corresponding document. The *codes* and the documents can be found in `DocumentsMain.csv`. A description of the column names Q1 – Q12 can be found in table 3. Table 4 is a description of all remaining columns of the file.

Table 3. Columns Q1 – Q12 of DatasetMain.csv. For the questions using the likert scale the encoding is: 1=Stronlgy disagree, 2=Disagree, 3=Neither agree nor disagree, 4=Agree, 5=Strongly agree.

Question ID	Question	Possible answers
Q1	What is your age?	Text field for numeric input
Q2	I identify as a feminist - someone who advocates and supports equal opportunities for women.	Likert scale
Q3	How would you describe your political view?	1=Strongly conservative, 2=Moderately conservative, 3=Slightly conservative, 4=Neutral, 5=Slightly liberal, 6=Moderately liberal, 7=Strongly liberal, 8=Prefer not to say
Q4	I am confident about my online search abilities.	Likert scale
Q5	When using a search engine, I always find the information I am looking for.	Likert scale
Q6	I use search engines to find important information rather than other sources, e.g., books, newspapers.	Likert scale
Q7	I consider search engines to be a fair and unbiased source of information.	Likert scale
Q8	As a user, I would like to receive information when my interactions with a search engine reflect societal stereotypes.	Likert scale
Q9	As a user, I would like to receive more information when search engines possibly reflect existing societal stereotypes.	Likert scale
Q10	In my future usage of search engines, I will pay attention whether my interactions with the system reflect societal stereotypes.	Likert scale
Q11	What gender do you identify with?	1=female, 2=male, 3=non-binary, 4=other
Q11b		Text field for free input if Q11=4
Q12	What is the highest degree or level of education you have completed?	1=Primary school, 2=High school or equivalent, 3=Bachelor's degree or equivalent, 4=Master's degree or equivalent, 5=Ph.D. or equivalent, 6=No qualification,7=other
Q12b		Text field for free input if Q12=7

Table 4. Column names of DatasetMain.csv except for the *codes* and Q1 – Q10. Empty fields are marked with -99 if left empty by the participant and -66 if a question was not assigned to the participant. For the questions using the likert scale the encoding is: 1=Stronlgy disagree, 2=Disagree, 3=Neither agree nor disagree, 4=Agree, 5=Strongly agree.

Column name	Description of column content
ParticipantId	Unique ID of participant (from prolific.co)
TimeTaken	Time taken for the survey completion in seconds
Age	Age declared to prolific.co
Sex	Sex declared to prolific.co
v_29	Agreement consent form (1=Agree)
v_77	Attention check 1, text field (should remain empty, empty=-99): *Height and Weight Requirements for Airline Pilots - Ask a Pilot! This is an attention check. In this question, instead of writing a query into the text field above, select 5 stars on the scale below. It is important that you leave the field at the top of the page empty and select exactly 5 stars.*
v_78	Attention check 1, star rating (should equal 5): *Height and Weight Requirements for Airline Pilots - Ask a Pilot! This is an attention check. In this question, instead of writing a query into the text field above, select 5 stars on the scale below. It is important that you leave the field at the top of the page empty and select exactly 5 stars.*
v_80	Attention check 2 (should be 1): *I have never used the internet.*(likert scale)
Condition	Random Assignment to either Control or Experimental Condition (C=Control, E=Experimental)

3.5 Description of `DatasetMainLabeled.csv`

Table 5 is a description of the file `DatasetMainLabeled.csv` which contains the labelled results from the main study.

Table 5. Description of DatasetPilotLabeled.csv. The file contains the labelled results from the pilot study.

Column name	Description of column content
ParticipantId	Unique ID of participant (from prolific.co)
ParticipantGender	Gender of the participant (f=female, m=male)
Condition	Random Assignment to either Control or Experimental Condition (C=Control, E=Experimental)
Domain	Domain the document belongs to (0=Appearance, 1=Career, 2=Child Care, 5=Physical Capabilities)
DocumentId	DocumentId is a unique ID for each document
GenderIndication	Gender indication of the document variation (f=female, m=male, n=neutral)
ExpStereotype	Expected Stereotype depending on the domain a document belongs to (f=female, m=male, u=undefined)
Query	Query formulated by the participant
Label	Label denoting the gender mentions in the query (f=female, m=male, n=none, r=rejected query)

REFERENCES

[1] Klara Krieg, Emilia Parada-Cabaleiro, Gertraud Medicus, Oleg Lesota, Markus Schedl, and Navid Rekabsaz. 2022. Grep-BiasIR: A Dataset for Investigating Gender Representation-Bias in Information Retrieval Results. https://doi.org/10.48550/ARXIV.2201.07754