# Social Information Retrieval Experiment & Exercise

Fact Pack for Homework Assignment #4

Christoph Fuchs

2015-05-20

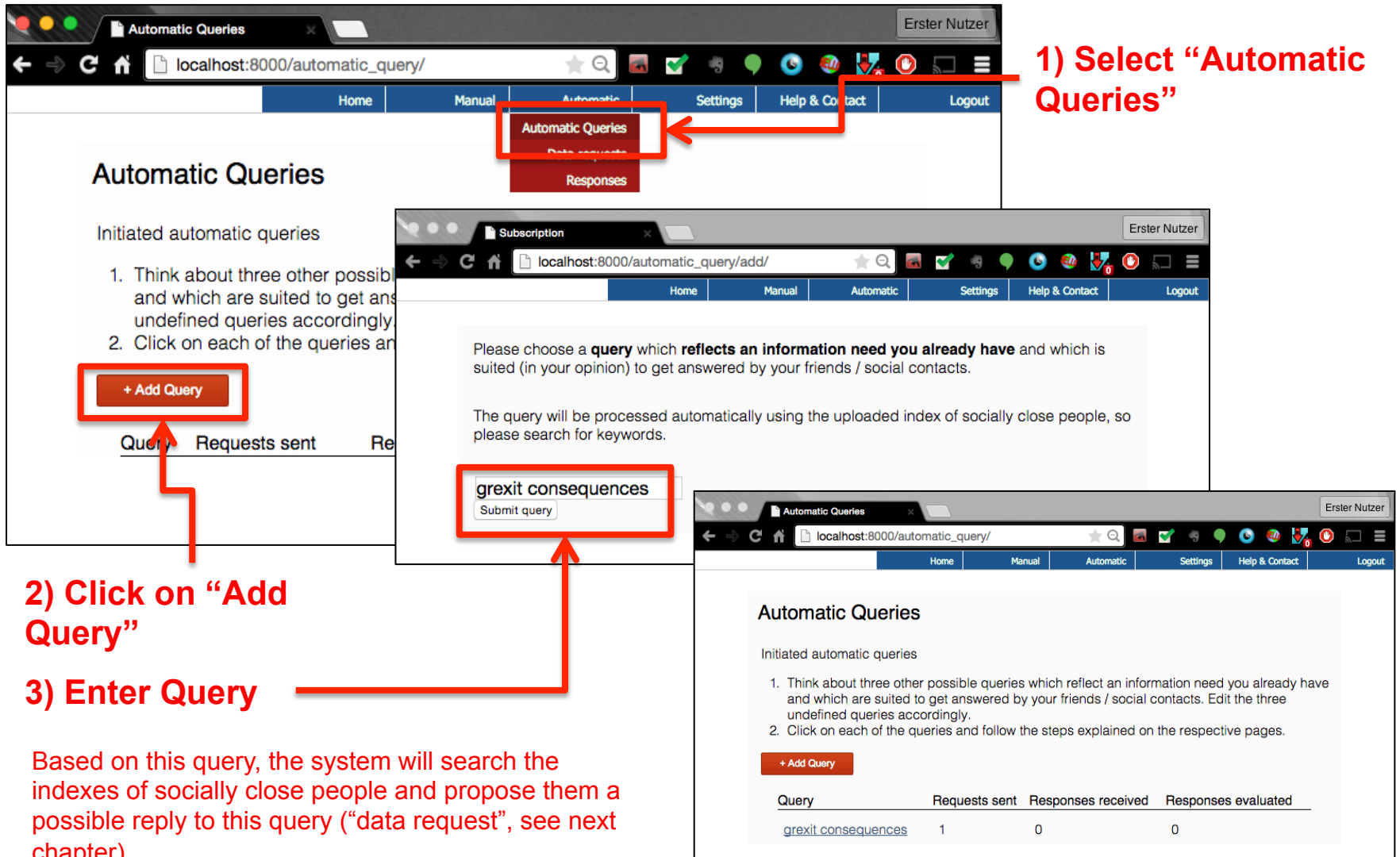*Source: https://vmschlichter24.informatik.tu-muenchen.de/ha4/*



Covered area in this document

# Homework Assignment #4

Due: **17.06. 23.59 CEST**

1. Start automatic query process - see fact pack for further details (add 6 queries to the automatic mode)
2. Maintain manual process (evaluate replies)
3. Analyze "our" anonymized social network
    ○ Create a Python script (using the igraph library) to
        ▪ Identify number of components
        ▪ Apply two different clustering algorithms
        ▪ Calculate diameter, density, clustering coefficient of the network
    ○ Comment on the results in a report (~50-100 words): Which results are surprising, which not? Give reasons for your answers.
4. Upload the Python script and your file with the comments

# Add Queries in Automatic Mode

# 1) Add a Query in Automatic Mode



**1) Select "Automatic Queries"**
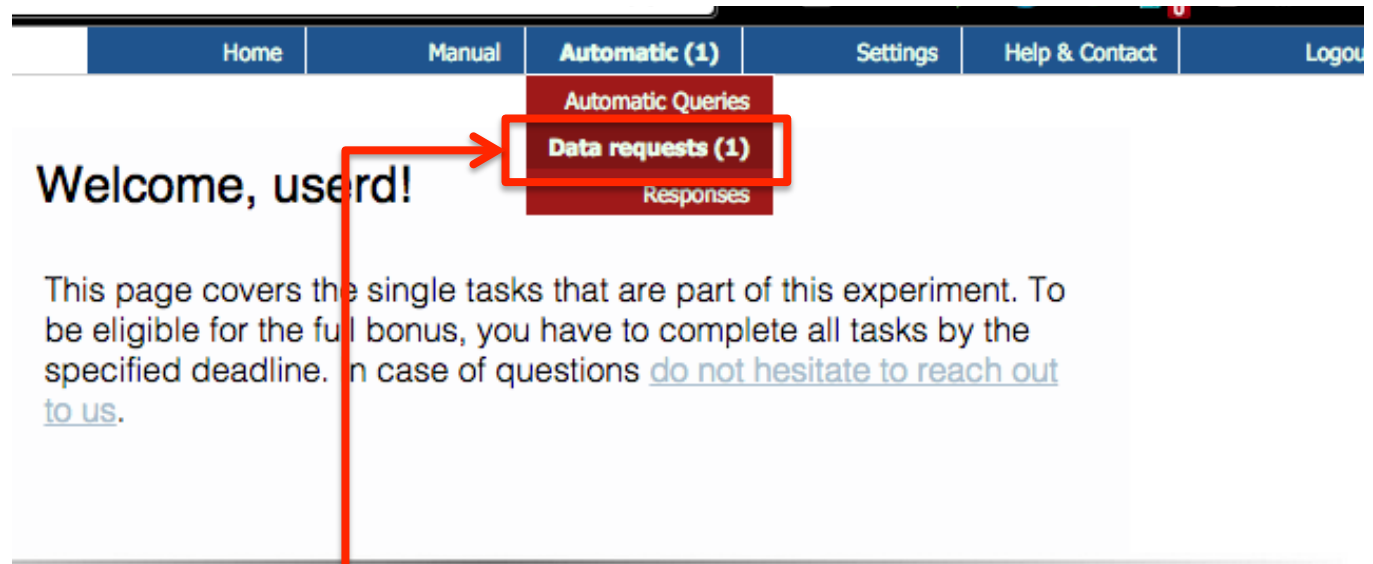
**2) Click on "Add Query"**

**3) Enter Query**

Based on this query, the system will search the indexes of socially close people and propose them a possible reply to this query ("data request", see next chapter)
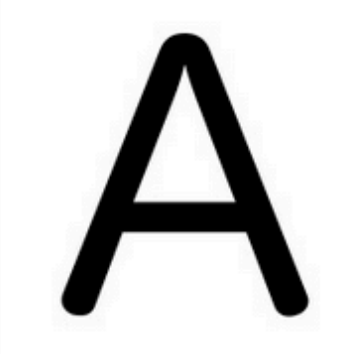
# Reply to data requests in Automatic Mode

# 1) Visit the web system
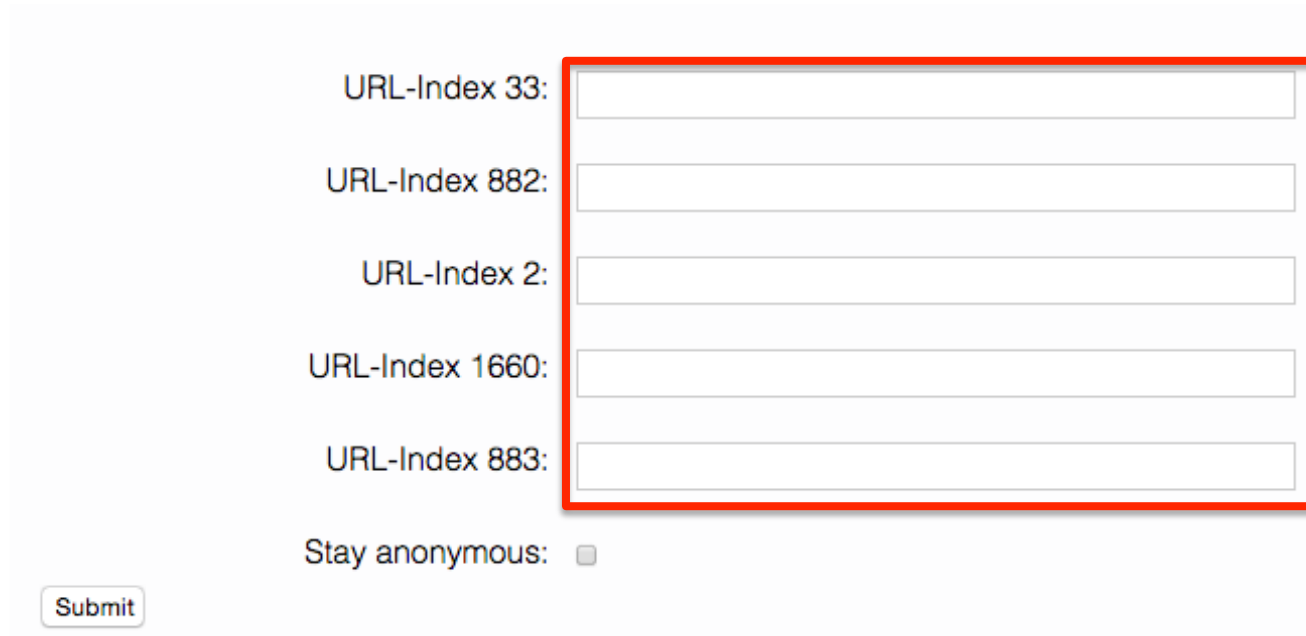


**Select "Data requests" in the "Automatic" folder**

**2) Select the data request to reply to**

# 3) Enter the URLs of your history (1/2)

URL-Index 33:

URL-Index 882:
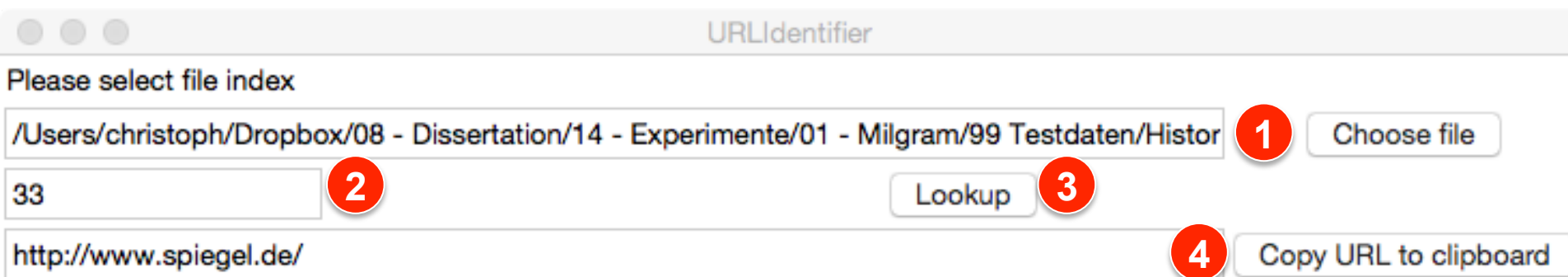
URL-Index 2:

URL-Index 1660:

URL-Index 883:

Stay anonymous: ☐

Submit

**Fill those fields using URLIdentifier (see next slide)**

**TΛΠ**

## 3) Enter the URLs of your history (2/2)

- Run the URLIdentifier tool from the software kit



URLIdentifier

Please select file index

/Users/christoph/Dropbox/08 - Dissertation/14 - Experimente/01 - Milgram/99 Testdaten/Histor  **1**  Choose file

33  **2**  Lookup  **3**

http://www.spiegel.de/  **4**  Copy URL to clipboard

**1** Choose the URL_Mapping.pkl file generated while calculating your topic models (is in the same directory as the lda_model.model file)

**2** **3** Enter the requested index (taken from the web interface), click on "Lookup" (or press return)

**4** Copy resulting URL to clipboard

- Paste URL in respective form in web system (see previous slide)

# Evaluate Replies to Automatic Queries

# 1) Select Response to assess



1) Select "Respones" from the "Automatic" menu

2) Select "View & evaluate" to see (and assess) the response

## 2) Fill Evaluation Form



1) **Review provided result**

2) **Answer questions (e.g. by using sliders)**

# PREPARATION – NOTES & REMARKS (1/2)

**We take Data Privacy seriously**: We don't do anything without your permission & we are transparent on what we do:

1. **We don't know the names of your Facebook friends**: We only ask you to upload the hashed version of the network file[1]

2. **None of your data** (evaluation of ties, names, products, etc.) **will ever be related to you**

3. **We don't get to know your browser history** – the uploaded topic models only contain document IDs and word vectors

1 The social network built from all individual ego-networks will get hashed again and randomly changed (edges and nodes changed/removed) before used in exercise to ensure that it is not possible to identify nodes

# PREPARATION – NOTES & REMARKS (2/2)

4. The data will only be used to **conduct sound scientific research** (and student exercises)
5. You will use a (**obfuscated**) version of the dataset to **solve homework assignments**
6. Your **name, matriculation number, email address**, etc. **is only used to grade the homework** – it **will be deleted afterwards**
7. We will **never give** the data to anyone
8. Publications based on this dataset will only contain **highly aggregated information**
9. We will **hash the Amazon products after the experiment** and delete all copies of the original dataset