**General framework**

- **Input**
  - ClueWeb 12
  - Profiles and attractions
    - 5.5 / 27.3 TB
    - 33,447 WARC files
    - 733,019,372 documents
  - Profiles and attractions files
    - 562 user profiles
    - 50 attractions in Philadelphia, PA
    - 28,100 ratings to description and site

- **Filtering**
  - Exact mention of the context
  - Format: <City, ST>
  - Exclude non-related sentences mentioning the city in other state
  - Found 13,548,982 documents

- **Profile generation**
  - Based on attraction’s description using the user’s rating

- **Representation and transformation**
  - Efficient in terms of size: from 918 GB to 40 GB
  - Processing speed

- **Discussion**
  - **Effect of filtering**
    - Geographically appropriate documents are missing
    - Based on attraction’s description using the user’s rating
    - A sub-collection provided by the organisers (given) has a higher coverage of geographically appropriate documents than our filtered sub-collection
    - **Effect of personalisation**
      - Not easy to compare: low coverage of relevance assessments
      - Method | MRR | MRR$_{5}$ | AP@5 | AP$_{5}$
      - IBCosTop1 | 0.0559 | 0.0745 | 0.0587 | 0.2202
      - IBCosTop1 + 5NN text cos | 0.0455 | 0.0562 | 0.0330 | 0.1486
      - IBCosTop1 + 5NN text Jacc | 0.0433 | 0.0531 | 0.0330 | 0.1294
      - IBCosTop1 + 5NN rating cos | 0.0429 | 0.0533 | 0.0349 | 0.1477
      - IBCosTop1 + 5NN rating Pearson | 0.0406 | 0.0508 | 0.0358 | 0.1560
      - Classifier + 5NN text cos | 0.0045 | 0.0123 | 0.0045 | 0.0290
      - Classifier + 5NN text Jacc | 0.0045 | 0.0123 | 0.0045 | 0.0290
      - Classifier + 5NN rating Pearson | 0.0045 | 0.0067 | 0.0018 | 0.0253
      - Positive profile | 0.0396 | 0.0558 | 0.0359 | 0.1498
      - Negative profile | 0.0045 | 0.0045 | 0.0009 | 0.0152
      - Positive + 5NN text cos | 0.0045 | 0.0072 | 0.0041 | 0.1399

- **Effect of evaluation**
  - We found a bias towards methods using documents from the Open Web
  - Method | Open Web | ClueWeb12 | Documents in the intersection
  - IBCosTop1 | 0.0009 | 0.045 | 4.000
  - IBCosTop1 + 5NN text cos | 0.742 | 0.650 | 2.767
  - Classifier + 5NN text cos | 0.009 | 0.751 | 2.121
  - Classifier + 5NN text Jacc | 0.431 | 0.140 | 1.422
  - Classifier + 5NN rating Pearson | 0.418 | 0.702 | 1.879
  - Classifier + 5NN rating Pearson | 0.083 | 0.452 | 1.566