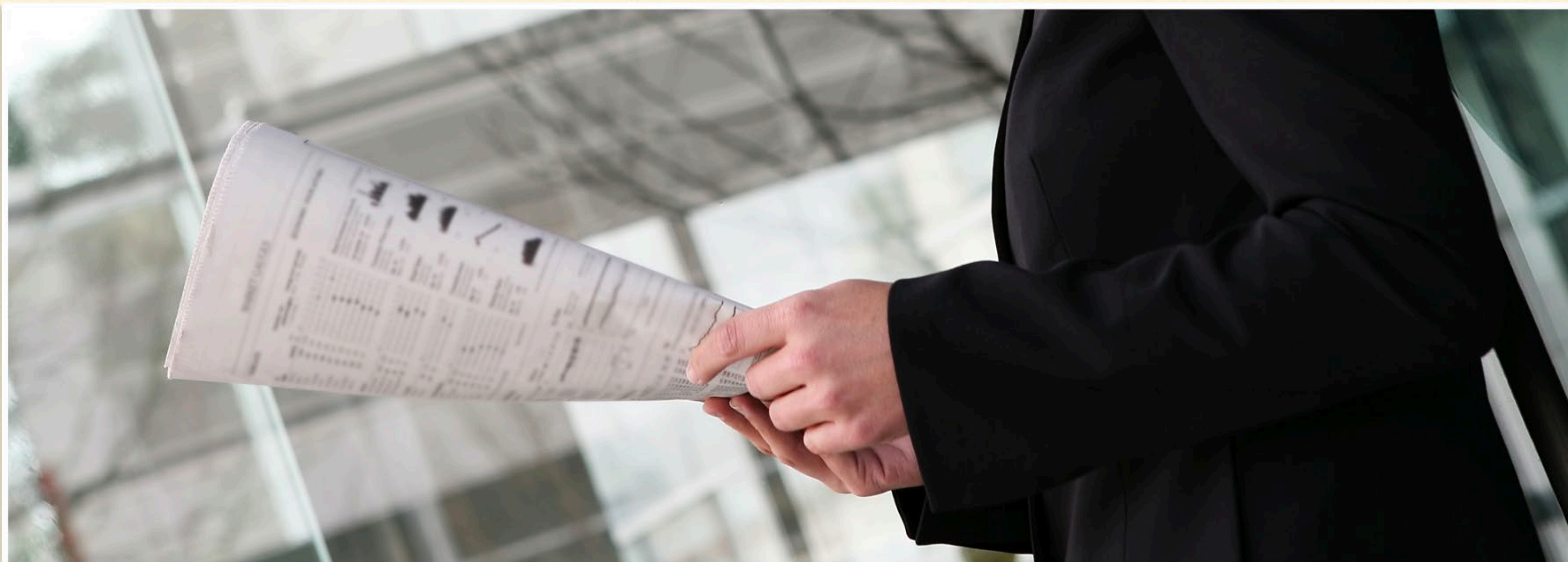




Using MTurk to Improve Content Analyses

Content Analysis PDW

Tim Hubbard, Ph.D.



UNIVERSITY OF
NOTRE DAME

Mendoza College of Business

You just collected thousands of media articles about a sample of firms...

How do you feel confident they are actually about the firm?

Why Should We Care?



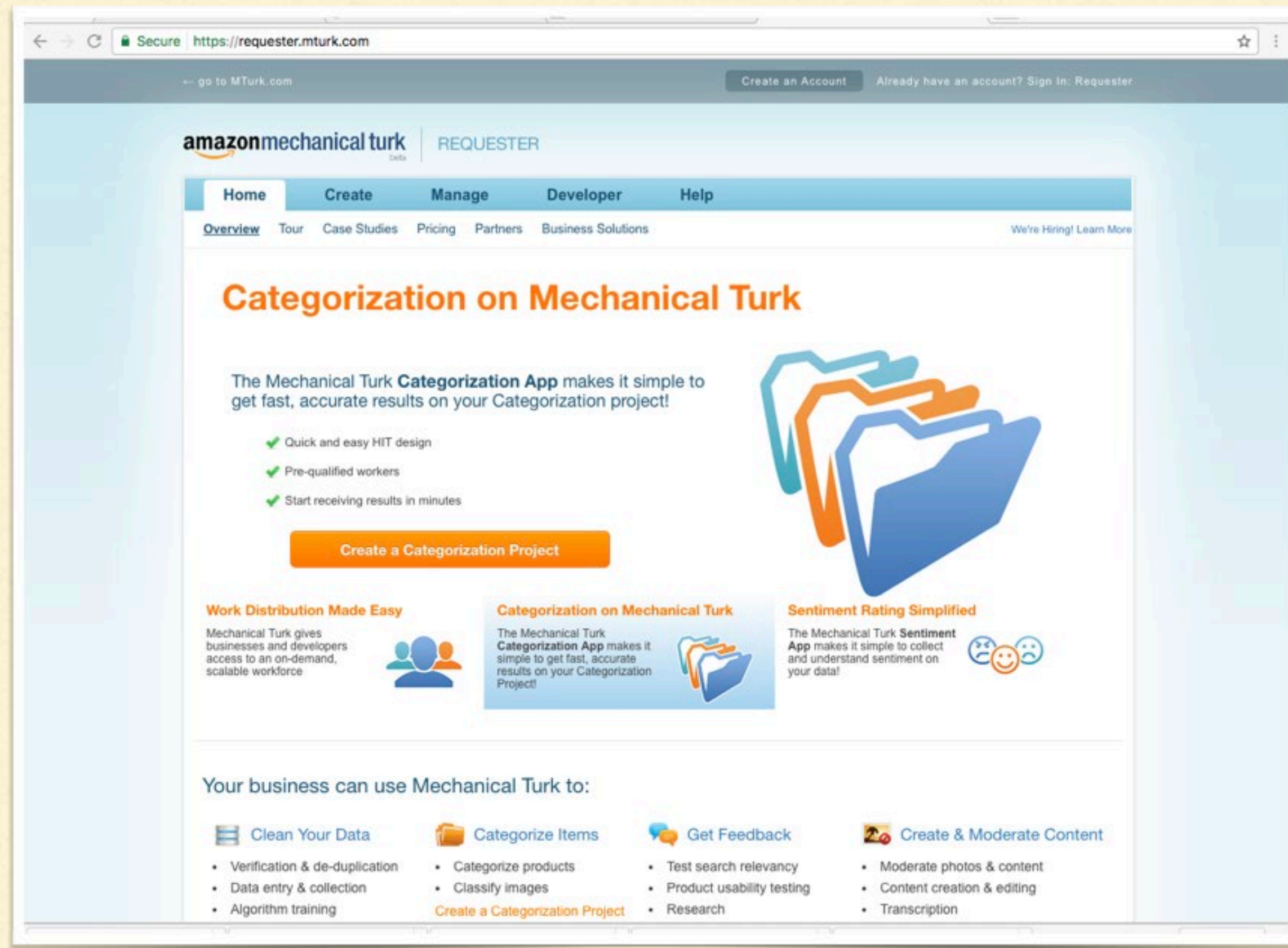
- Stronger connection between construct and proxy
- Reduces measurement error
- Increases the likelihood of finding results
- Thoroughness of independent ratings are more valid and easy to report

Our Situation

- Study of celebrity startup firms
- Reviewer requested we collect more media data
- Needed to analyze general media's tenor, volume, and non-conformity of sample firms
- Needed to do it thoroughly—and fast

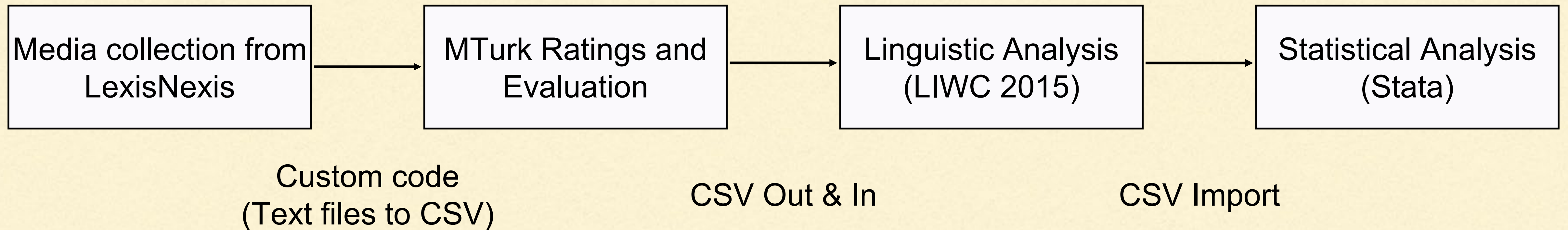


Amazon's MTurk Helped

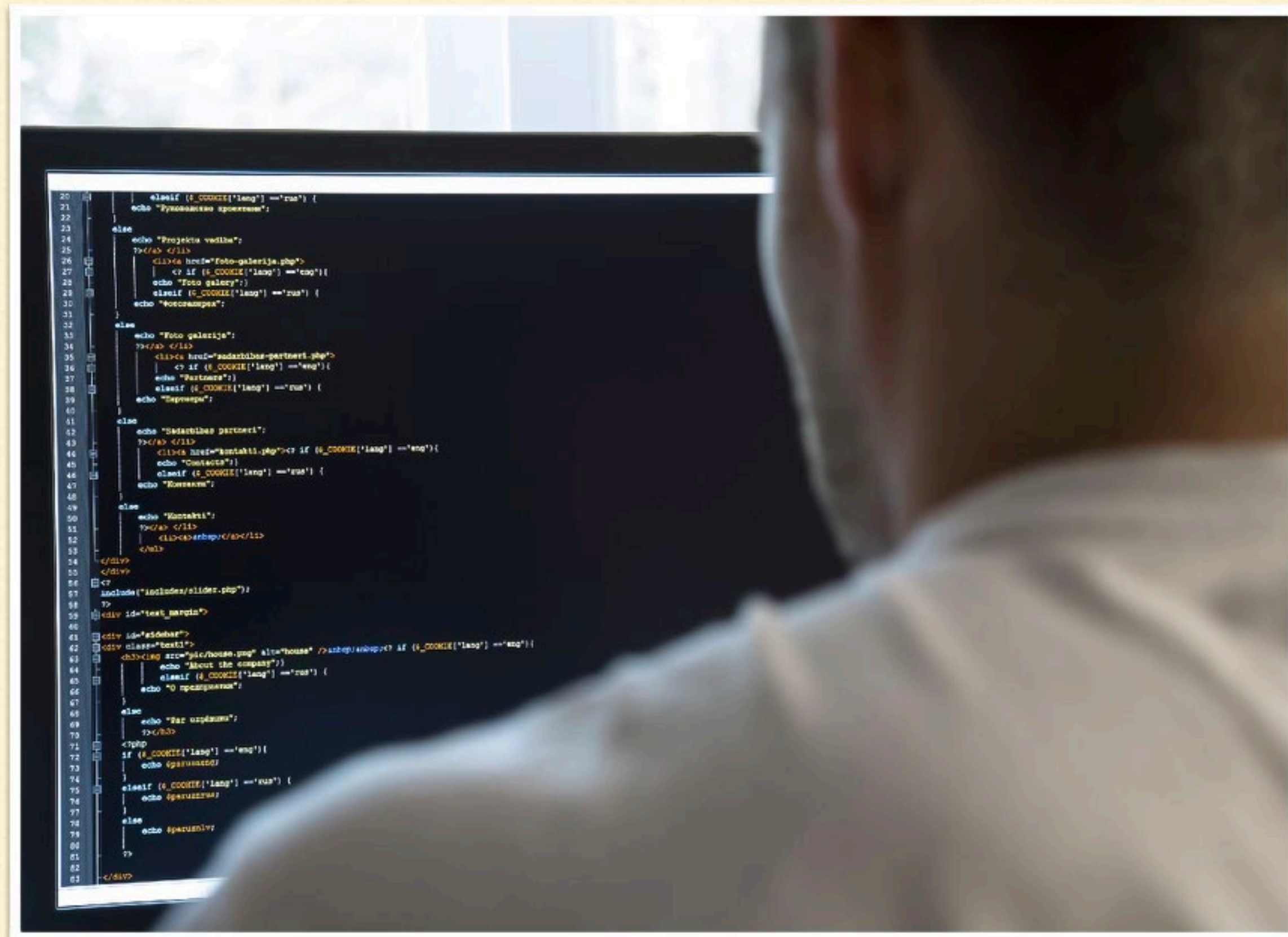


- Human coding
- Gives access to multiple individual raters per article
- Very fast
- Imports data using CSV files
- Easy payment

General Process



Parameters



- 6,260 articles from LexisNexis
- Presented only the first page of the articles
- 2 raters per article (12,520 ratings)
 - Disagreement or unclear, added one more (795)
- 7 hours per run
- \$0.08/rating

Frame Height

550

Height in pixels of the frame your HIT will be displayed in to Workers. Adjust the height appropriately to minimize scrolling for Workers.

Format

Font

A

*I*_x

≡

≡

≡

≡

≡

≡

↩

↪

🔗

🔗

📄

Source

Instructions (Click to expand)

Is \${company} the primary focus of this article?

Value	Guidance
Yes	Select this if the article primarily about the company \${company}.
Unclear	Select this if it is unclear to you whether the article is primarily about on \${company}.
No	Select this if the article is not primarily about the \${company}. For example, it might be an article about Microsoft that mentions \${company}.

Article Title: \${title}

Article Body: \${body}

Company: \${company}

Reference: \${identifier}

Company Ticker: \${ticker}

Is \${company} the primary focus of the article?

Yes

Unclear

No

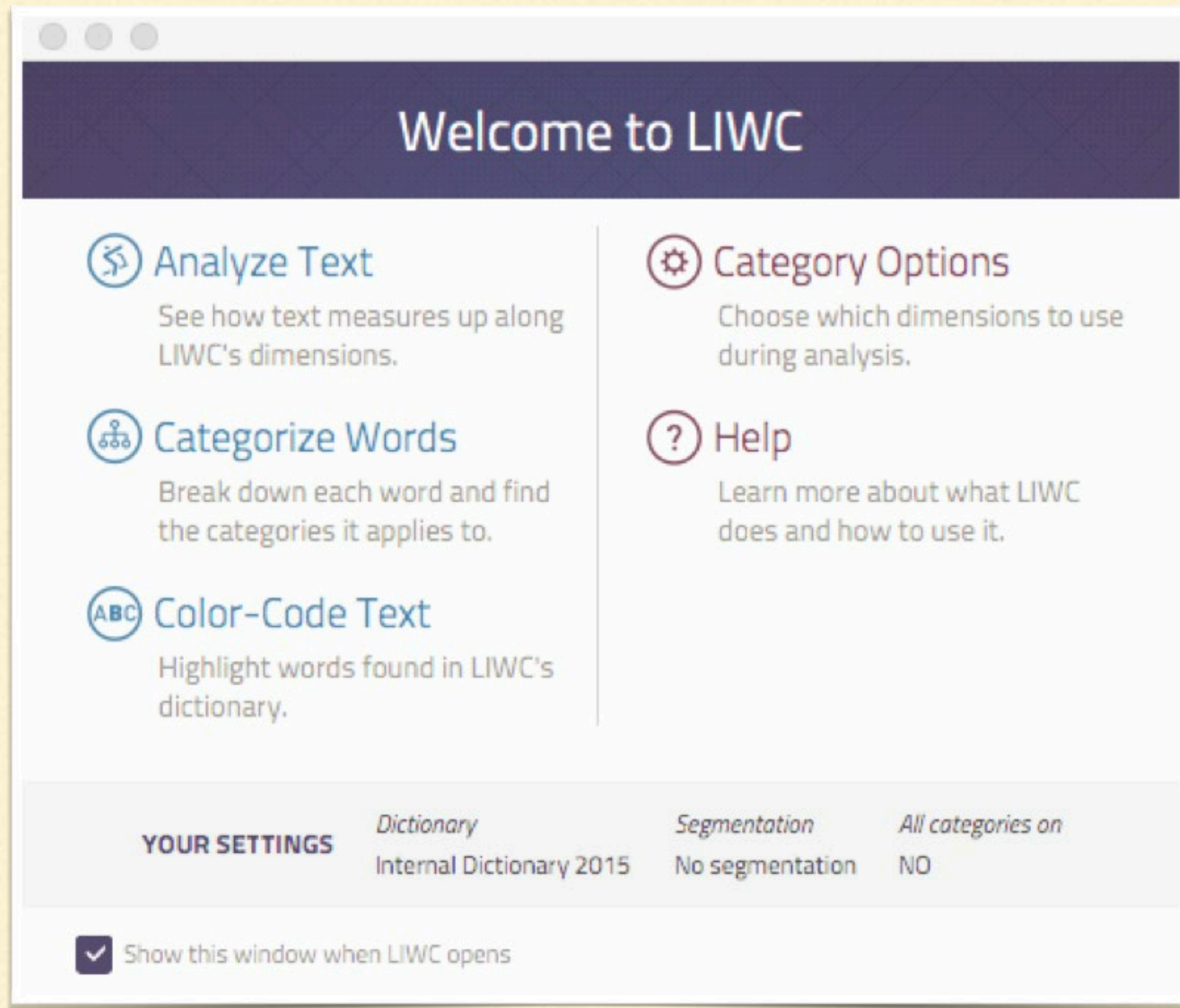
body

Moving from LexisNexis to MTurk

```
7 #what uses rtf files. based on the file extension, the parsing is handled a little differently. A CSV file is also
8 #built that contains a row for each article.
9
10 ##Note! To use this script, you must have coreutils installed via brew
11 # brew install coreutils
12
13 ##Make sure we received the right number of arguments:
14 if [ $# -ne 1 ]
15 then
16     echo "Usage: ./mediaParser.sh <Directory To Parse>"
17     exit 0
18 fi
19
20 function getFileName {
21     local filename=$(basename "$1")
22     echo "${filename%.*}"
23 }
24
25 function getFileExtension {
26     local filename=$(basename "$1")
27     echo "${filename##*.}"
28 }
29
30 #Setup some easier to remember variable names:
31 DIRECTORY_TO_PARSE=$1
32
33 rm ./toMTURK.csv
34 rm -rf ./processed-articles
35 mkdir processed-articles
36
37 for FILE_TO_PARSE in ./${DIRECTORY_TO_PARSE}/*; do
38     FILE_NAME=$(getFileName $FILE_TO_PARSE)
39     FILE_EXTENSION=$(getFileExtension $FILE_TO_PARSE)
40     #Parse the file:
41     if [ $FILE_EXTENSION = "rtf" ]
42     then
43         gcsplit -s --elide-empty-files --digits=2 --prefix="./processed-articles/pre-$FILE_NAME-$FILE_EXTENSION-" $FILE_TO_PARSE "/"
44
45         #Remove the last file because it does not have anything interesting in it:
46         ls -t ./processed-articles/pre-${FILE_NAME}* | tail -n 1 | xargs rm -f
47     else
48         gcsplit -s --elide-empty-files --digits=2 --prefix="./processed-articles/$FILE_NAME-$FILE_EXTENSION-" $FILE_TO_PARSE "/DOCUMENT"
49
50         #Remove the first file because it will be empty:
51         ls -t ./processed-articles/${FILE_NAME}* | head -n 1 | xargs rm -f
52     fi
53 done
54
55 echo 'identifier,title,body' > ./toMTURK.csv
```

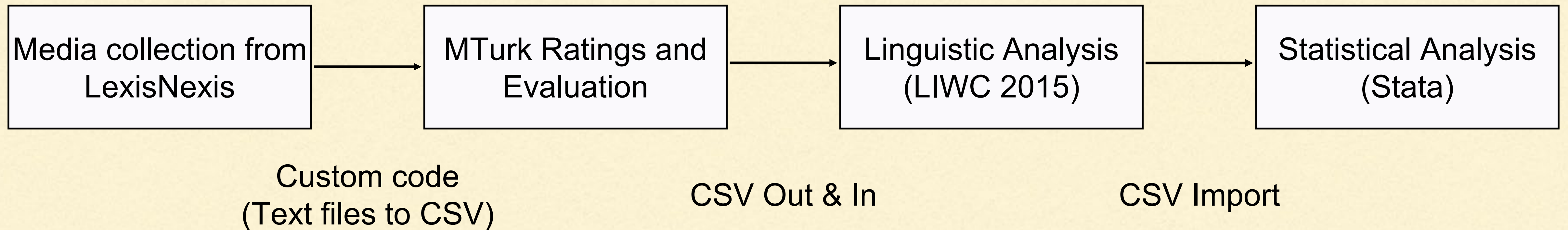
- Custom developed software to move from LexisNexis text files to a fully populated CSV file
- That file can then go up to MTurk
- Developed by Rampant Strategy OÜ

LIWC 2015



- Can import CSV files now
 - Straight from MTurk
- Can export to CSV which is easy to import into Stata

General Process



Questions & Discussion

thubbard@nd.edu

