Measuring occupations: respondent’s self-identification from a large database

SPECIAL SESSION: Synergies for Europe’s Research Infrastructures in the Social Sciences and Official Statistics (SERISS)

THU 2 June 2016

Kea Tijdens
University of Amsterdam, NL
k.g.tijdens@uva.nl
The national stocks of job titles are ...

• large >> 10,000’s of job titles in any national labour force
• unstructured >> vague boundaries between job titles
• unlimited >> no fixed list, many entries and exits over time

The challenge

• to classify job titles into ISCO-08 classification of occupations
• & ... to do so consistently across countries

Occupational titles vs job titles

• job titles: within organisational context
• occupational titles: beyond organisational context
‘What is your occupation?’

Open-ended questions: textbox
- Textbox: predominantly used in surveys
- Office coding needed: expensive and time-consuming, though increasingly high quality coding software and auto-coders
- Coding problems: vague or highly aggregated titles

Closed questions: dictionaries
- Brief list (max 10 entries): predominantly used in postal surveys >> aggregation bias
- Showcard (max 50 entries): predominantly used in face-to-face surveys
- Dictionary (unlimited number of entries): respondents self-select their occupation from a list of occupations
Closed question: the example of WageIndicator web survey

WageIndicator websites
- In 2001: website with and wage content - started in the Netherlands
- Today: web portal with websites national websites in 89 countries, all in national language(s)
- 2015: 32 million of visitors, most through search engines

WageIndicator multilingual, continuous web survey
- All websites invite visitors to complete
  - a long salary survey in return to free information provided, with lottery incentive
  - or a mini-survey to get a salary indication

“What is your occupation?”
- Closed survey question (coding too expensive, particularly with more languages included)
- Respondents self-identify their occupation through
  - a search tree (IPod menu) and an autosuggest box (Google search type)
- Dictionary: a multi-lingual database of occupations, all coded 4 digit ISCO-08

History of this closed survey question
- 2001-’05 700 occupations in 1 language with 2-level search tree >> one page per level
- 2006-’09 1,100 occupations with 8 languages with 3-level search tree >> one page per level
- 2009-’15 1,600 occupations with 30 languages with 3-level tree on one page + autosuggest
- 20015-.. Database in 41 languages available on an API (Application Programming Interface)
Search tree (left), autosuggest (right)
SERISS project (2015-2019)

Extend the dictionary of occupations
- To 99 countries with 47 languages
- To serve self-identification through search tree and through autosuggest
- Ensure that all occupational titles are well coded in ISCO-08

Make database available for survey holders
- Program an API for use in web surveys on desktop, tablet, smartphone
- Program an interface for use in CAPI surveys
- Make database downloadable in excel
- Availability: till end SERISS free of charge

Develop an occupation – industry prediction
- Depending on ticked occupation, a limited set of industries is shown for the survey question ‘In which industry do you work?’
- Aiming to reduce respondents’ time
Search tree vs autosuggest

Meta Data occupation API

- 23 – 30 May 2016
- 12 436 records from clicks and autosuggest >> 1990 respondents

Use of search tree versus autosuggest

- More than three in four uses the search tree (78%)
- More than one in five uses the text box (22%)
- Drop out rate (5%) [Note: drop out is common in this web survey]

Do not quote, because not controlled for mobile use

- Mobile users use only search tree

<table>
<thead>
<tr>
<th></th>
<th>complete</th>
<th>dropout</th>
<th>Total</th>
<th>%</th>
<th>% drop out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search tree</td>
<td>1477</td>
<td>70</td>
<td>1547</td>
<td>78%</td>
<td>5%</td>
</tr>
<tr>
<td>Autosuggest</td>
<td>409</td>
<td>34</td>
<td>443</td>
<td>22%</td>
<td>8%</td>
</tr>
<tr>
<td>Total</td>
<td>1886</td>
<td>104</td>
<td>1990</td>
<td>100%</td>
<td>5%</td>
</tr>
<tr>
<td>% complete</td>
<td>95%</td>
<td>5%</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Do respondents go back and forth in the search tree?

- 1547 respondents started the search tree
- 54% found their occupation in three clicks
- 14% went back and forth one time
- 29% went back and forth more than one time

<table>
<thead>
<tr>
<th>back and forth</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>drop out</td>
<td>2.3</td>
</tr>
<tr>
<td>0</td>
<td>54.5</td>
</tr>
<tr>
<td>1</td>
<td>14.0</td>
</tr>
<tr>
<td>2</td>
<td>7.6</td>
</tr>
<tr>
<td>3</td>
<td>6.3</td>
</tr>
<tr>
<td>4</td>
<td>3.4</td>
</tr>
<tr>
<td>5</td>
<td>2.0</td>
</tr>
<tr>
<td>&gt;5</td>
<td>9.8</td>
</tr>
<tr>
<td>total</td>
<td>99.9</td>
</tr>
</tbody>
</table>
Response time

Response time in seconds

- After controlling for outliers (min 1 second, max 360 seconds) and for drop outs, response times of 1843 respondents were analysed.
- Mean response time larger for autosuggest than for search tree (48 versus 44 seconds).
- Median response time larger for search tree than for autosuggest (26 versus 18 seconds).

<table>
<thead>
<tr>
<th>Seconds</th>
<th>Minimum</th>
<th>Median</th>
<th>Mean</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search tree</td>
<td>1</td>
<td>26</td>
<td>44</td>
<td>352</td>
</tr>
<tr>
<td>Autosuggest</td>
<td>1</td>
<td>18</td>
<td>48</td>
<td>357</td>
</tr>
</tbody>
</table>
Selected occupations ISCO-08

• The 1888 respondents selected 677 unique titles from the list of 1,600 titles
• Graph shows the distribution of sample for to 1-digit ISCO-08 classification
The end ..... 

Thank you for your attention 😊😊

Please try the demo: 
http://tmt.centerdata.nl/jobcoder_demo/

Questions?
k.g.tijdens@uva.nl