Self-identification of occupation in web-surveys: Respondents’ choice between autosuggest and search tree

2017 CSDI Workshop
Kea Tijdens,
University of Amsterdam/AIAS
17 – 03 - 2017
• **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 for multi-country surveys**
  - 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 survey questions: textbox**
  - Textbox: used in most surveys
  - Office coding needed: expensive and time-consuming, though increasingly high quality coding software and auto-coders
  - Coding problems: vague job titles, highly aggregated titles, company-specific titles & typos

• **Closed survey 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
The WageIndicator web survey in 91 countries

**WageIndicator websites**
- In 2001 in NL: website with job related content, salaries per occ.
- Today: web portal with national websites in 91 countries, all in national language(s)
- 2016: 40 million of visitors, most through search engines

**Multilingual, continuous web survey**
- All websites invite visitors to complete (>> opt-in survey)
  - a long salary survey with lottery incentive
  - or a mini-survey to get a salary indication
- Large numbers of survey respondents
- Over the past 15 years tremendous technological progress
- Survey frequently used for micro-targeting specific respondents through social media and through face-to-face surveys with app
Closed survey question: “What is your occupation?”

Self-identification: search tree or autosuggest

- Closed survey question, because coding was too expensive
- Respondents self-identify their occupation through
  - a search tree (IPod menu) or
  - an autosuggest box (Google search type)
- Dictionary: multi-lingual database of occupations coded ISCO08

History of the closed question & database

<table>
<thead>
<tr>
<th>Years</th>
<th># occ’s</th>
<th># lang.</th>
<th>search tree</th>
<th>auto s.</th>
<th>API</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001-’05</td>
<td>700</td>
<td>1</td>
<td>2-level, 1 page p/lvl</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2006-’09</td>
<td>1,100</td>
<td>8</td>
<td>3-levels, 1 page p/lvl</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2009-’12</td>
<td>1,600</td>
<td>30</td>
<td>3-levels, 1 page p/tree</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2012-’15</td>
<td>1,600</td>
<td>30</td>
<td>3-levels, 1 page p/tree</td>
<td>yes</td>
<td>-</td>
</tr>
<tr>
<td>2015-’17</td>
<td>1,600</td>
<td>43</td>
<td>3-levels, 1 page p/tree</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
Search tree (left), autosuggest (right)
本题是必答题
您的职业是什么？
如果您的职业不在列表中，请选择与您职业最相近的一个。

无线电

无线电通信线路或光缆装配工或修理工
无线电通信工程师
无线电，电视或其他媒体播音员
无线电设备操作员
船只或无线电通信官员

手机设备装配工或修理工
无线电通信工程师
无线电通信线路或光缆装配工或修理工
**Search tree vs autosuggest**

- **Meta Data occupation API in WageIndicator web survey**
  - API data 26 June – 3 Nov 2016, selection nl_NL
  - API data registers each click and each character in autosuggest
  - 18,448 records of 2,994 respondents

- **Descriptives users**
  - One in four uses a mobile device (25.6%)
  - Two in three start using the search tree (66.9%)
  - Mobile users use less often search tree (Chisq 57.37, sign .000)

<table>
<thead>
<tr>
<th></th>
<th>No mobile</th>
<th>Mobile</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search tree</td>
<td>70.7%</td>
<td>55.8%</td>
<td>66.9%</td>
</tr>
<tr>
<td>Autosuggest</td>
<td>29.3%</td>
<td>44.2%</td>
<td>33.1%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
Do respondents go back and forth in the search tree?

- 2004 respondents started the search tree
- 3% dropped out with one or two clicks
- 54% found their occupation in three clicks
- 15% went back and forth one time
- 28% went back and forth more than one time

<table>
<thead>
<tr>
<th>clicks</th>
<th>0</th>
<th>b&amp;f1</th>
<th>b&amp;f&gt;1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>2%</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>1%</td>
</tr>
<tr>
<td>3</td>
<td>1081</td>
<td>0</td>
<td>0</td>
<td>54%</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>301</td>
<td>0</td>
<td>15%</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>174</td>
<td>9%</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>94</td>
<td>5%</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>77</td>
<td>4%</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>2%</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>1%</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>152</td>
<td>6%</td>
</tr>
</tbody>
</table>
Response time

- **Response time in seconds**
  - After controlling for drop outs and outliers (<1 - >360 sec.), response times of 2817 respondents were analysed.
  - Mean response time equal for autosuggest and search tree (39 sec.).
  - Median response time larger for search tree than for autosuggest (24 versus 16 sec.).

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Median</th>
<th>Mean</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search tree</td>
<td>1.7</td>
<td>23.6</td>
<td>39.1</td>
<td>347.0</td>
</tr>
<tr>
<td>Autosuggest</td>
<td>1.1</td>
<td>16.4</td>
<td>39.1</td>
<td>335.9</td>
</tr>
<tr>
<td>Total</td>
<td>1.1</td>
<td>21.6</td>
<td>39.1</td>
<td>347.0</td>
</tr>
</tbody>
</table>
Which respondents drop out?

- 5% of respondents drop out when they self-identify their occupation
- No sign difference between mobile and non-mobile users
- No sign difference between search tree and autosuggest users
- Search tree users are more likely to drop out after five clicks
- Autosuggest users are more likely to drop out after typing 20 characters
Selected occupations

- **Selected occupations ISCO-08**
  - The 2847 respondents selected 724 unique titles from the list of 1,600 titles (ISCO 5 digit)
  - Graph shows the distribution (%) for 1-digit ISCO-08
• **Extend the dictionary of occupations**
  – To 99 countries with 47 languages
  – To serve self-identification through search tree and autosuggest
  – Ensure that all occupational titles are well coded in ISCO-08

• **Make database available for survey holders**
  – Program API for 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
• **What is main activity of organisation where you work?**
  – Industry database with 320 entries
  – Coded 3 or 4 digit NACE2.0, crossovers to ISIC
  – Database & related survey question available in 47 languages

• **What is your highest education?**
  – >> See presentation of CAMCES database

• **What is your employment status?**
  – Survey questions & answers to identify employment status
  – Coded ICSE-93 (ILO classification) + prospective update
  – Coded ESEG-2014: Eurostat’s socio-economic classification (jointly with occupation code)
  – Survey questions & answers available in 47 languages
Available for survey holders?

- **Yes, survey module at** [http://surveycodings.org/](http://surveycodings.org/)
  - The full module consists of 32 survey questions
  - This includes questions for multi-country surveys: selection of language and country of residence
  - This includes questions for routing: year of birth, gender, in paid employment/unemployed
  - This includes questions in present and past tense for current or latest job
  - Survey holders can select questions they want to ask
  - SERISS provides free of charge until Jun2019

- **Workshop 4-5 September in Amsterdam**
  - Presentation of the module and demo
  - Presentations of databases
  - SERISS can provide travel&accommodation funding
Thank you for your attention 😊😊

Useful links

http://www.wageindicator.org/main

Please try the demo:

• http://tmt.centerdata.nl/jobcoder_demo/

Questions?

• k.g.tijdens@uva.nl