Can auto-text recognition software for coding injuries replace manual coding?

Findings from IDB/DISS in the Netherlands

Susanne Nijman, Consumer Safety Institute, Amsterdam, the Netherlands
(s.nijman@veiligheid.nl)

Birgitte Blatter, Consumer Safety Institute, Amsterdam, the Netherlands
What is VeiligheidNL?

Expertise center for safe behaviour in a safe environment

**WHAT:** our approach

Monitoring trends and causes of injuries

**WHY:** A safe home, travel and work environment, for everybody

by stimulating people in a positive way

**HOW:** we make sure that safe behaviour is natural

Development of interventions to stimulate safe behaviour

We share knowledge with several target groups
Dutch Injury Surveillance System (DISS)

- Since 1997
- Registration of injuries at EDs
- Representative, 14 of 87 EDs, 11% of visits
- Extrapolation to national figures
- Injuries/intoxications:
  - cause (home and leisure, work, sport, traffic, violence, self-harm) + reasons
- Annual upload to European Injury DataBase (IDB)
Background

- Until a few years ago all variables in IDB/DISS (such as injury mechanism, product involved, type of injury and body part involved) were coded manually by the staff of the ED.

- To reduce the administrative burden on EDs we developed a system for automatic text recognition software.
What do we ask from EDs?

Information that is already registered in their own Hospital Information System:
  • Personal characteristics: age, gender and postal code
  • Diagnosis
  • Hospitalization yes/no

Additional information in open text fields (integrated in their hospital information system) for all injuries and intoxications:
  • What happened?
  • Where/when did it happen?
  • What products were involved?
Aim

Can automatic text recognition software for coding injuries replace manual coding?

Examples:
1. Car driver, collision against tree, high speed accident
   • Desired output:
     • Injury mechanism: contact with object
     • Products involved: car, tree

2. Patient found intoxicated, used alcohol and speed
   • Desired output:
     • Injury mechanism: chemical mechanism
     • Products involved: alcohol, speed
Methods

• After assessment of several tools, we chose IBM SPSS Modeler.

• We taught the system from scratch how to code information on accidents and injuries.

• All possible words were classified into libraries and the system was taught how to interpret sentences.

• Comparison: IBM Modeler out – manual check out.
IBM Lotus Notes: Manuel check / corrections
Findings: 81% of injury mechanism coded correctly

<table>
<thead>
<tr>
<th>Injury mechanism autotext</th>
<th>Fall</th>
<th>Contact with object</th>
<th>Contact with person or animal</th>
<th>Foreign body</th>
<th>Threat to breathing</th>
<th>Chemical mechanism</th>
<th>Thermal mechanism</th>
<th>Electricity, radiation, explosion</th>
<th>Physical over-exertion</th>
<th>Unspecified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>42.500</td>
<td>1.009</td>
<td>289</td>
<td>12</td>
<td>8</td>
<td>33</td>
<td>16</td>
<td>3</td>
<td>216</td>
<td>147</td>
</tr>
<tr>
<td>Contact with object</td>
<td>1.979</td>
<td>13.370</td>
<td>182</td>
<td>104</td>
<td>8</td>
<td>44</td>
<td>27</td>
<td>9</td>
<td>85</td>
<td>140</td>
</tr>
<tr>
<td>Contact with person or animal</td>
<td>779</td>
<td>366</td>
<td>3.219</td>
<td>18</td>
<td>0</td>
<td>15</td>
<td>1</td>
<td>1</td>
<td>66</td>
<td>78</td>
</tr>
<tr>
<td>Foreign body</td>
<td>41</td>
<td>88</td>
<td>32</td>
<td>790</td>
<td>1</td>
<td>75</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>Threat to breathing</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>13</td>
<td>68</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Chemical mechanism</td>
<td>156</td>
<td>68</td>
<td>8</td>
<td>10</td>
<td>0</td>
<td>1.469</td>
<td>3</td>
<td>1</td>
<td>14</td>
<td>38</td>
</tr>
<tr>
<td>Thermal mechanism</td>
<td>35</td>
<td>30</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>85</td>
<td>401</td>
<td>9</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Electricity, radiation, explosion</td>
<td>5</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>69</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Physical over-exertion</td>
<td>182</td>
<td>64</td>
<td>45</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>868</td>
<td>67</td>
</tr>
<tr>
<td>Unspecified</td>
<td>2.544</td>
<td>2.711</td>
<td>864</td>
<td>167</td>
<td>19</td>
<td>1.221</td>
<td>89</td>
<td>14</td>
<td>726</td>
<td>3.720</td>
</tr>
<tr>
<td>Total</td>
<td>48.231</td>
<td>17.724</td>
<td>4.647</td>
<td>1.119</td>
<td>106</td>
<td>2.950</td>
<td>550</td>
<td>103</td>
<td>1.981</td>
<td>4.229</td>
</tr>
</tbody>
</table>

% true                      | 88%   | 75%                 | 69%                          | 71%          | 64%                | 50%              | 73%             | 67%                          | 44%                 | 88%        |
% unknown                   | 5%    | 15%                 | 19%                          | 15%          | 18%                | 41%              | 16%             | 14%                          | 37%                 |            |
% false                     | 7%    | 9%                  | 12%                          | 14%          | 18%                | 9%               | 11%             | 19%                          | 20%                 | 12%        |

Total true                  | 81%   |                     |                              |              |                    |                  |                 |                              |                     |            |

81% of the cases was coded correctly for injury mechanism by autotext recognition software
## Analysis

<table>
<thead>
<tr>
<th>Injury mechanism autotext</th>
<th>Fall</th>
<th>Contact with object</th>
<th>Contact with person or animal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>42.500</td>
<td>1.009</td>
<td>289</td>
</tr>
<tr>
<td>Contact with object</td>
<td>1.979</td>
<td>13.370</td>
<td>182</td>
</tr>
<tr>
<td>Contact with person or animal</td>
<td>779</td>
<td>366</td>
<td>3.219</td>
</tr>
<tr>
<td>Foreign body</td>
<td>41</td>
<td>88</td>
<td>32</td>
</tr>
<tr>
<td>Threat to breathing</td>
<td>10</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Chemical mechanism</td>
<td>156</td>
<td>68</td>
<td>8</td>
</tr>
</tbody>
</table>
Next steps

• Start with analysis of false and unknown cases (largest numbers and/or percentage false): manual text analysis, search for patterns

• Make adjustments in SPSS Modeler based on analysis of false and unknown cases

• Check if adjustments have the desired effect
Conclusions (1)

• It takes a lot of time to prepare proper text analysis

• For only products we have imported 9,000 terms (including synonyms and misspelled words)

• Words with double meaning cause difficulties (takes a lot of time)
Conclusions (2)

• First analysis showed: 81% of injury mechanism coded correctly

• We still check every record manually and correct if necessary

• The work that is done by the software makes coding at VeiligheidNL easier

• In the future we will be able to reduce the number of checks

• And most important: we have managed to reduce the administrative burden for ED’s!
Thanks for your attention!
Questions?

Susanne Nijman, Consumer Safety Institute, Amsterdam, the Netherlands
s.nijman@veiligheid.nl
Birgitte Blatter, Consumer Safety Institute, Amsterdam, the Netherlands
b.blatter@veiligheid.nl

www.veiligheid.nl/en