Towards an emerging animal disease surveillance system based on textual media analysis

S. Valentin, R. Lancelot, M. Roche
Monitoring of international animal health events

Official sources

Unofficial sources

Belarus independent media report ASF outbreaks

Independent Belarussian media reported that outbreaks of African Swine Fever (ASF) are regularly being found, but somehow do not get officially registered.

Pro-human rights news site Charter’97 described events around a recent incident in Gomel region, southern Belarus, on October 28. Many pigs had died at a farm in the Stolpnya village, 82 km north west from Gomel city, which forced municipal authorities to apply quarantine measures, including roadblocks and a ban on any transport of pork and pig products from the affected territory.
‘Platform for Automated extraction of animal Disease Information from the web’

- African swine fever
- Avian Influenza
- Foot-and-mouth disease
- Schmallenberg virus
- Bluetongue
What is done

(1) Data acquisition

(2) Data classification

(3) Information extraction

Relevant documents

Epidemiologic indicators

Location

Date

Disease

Host

Number of cases
Avian flu confirmed in 17 wild birds in Dorset

EC offers millions to help Poland's ASF-hit pig farmers

Schmallenberg virus: what you need to know
What is done

(1) Data acquisition

(2) Data classification

(3) Information extraction

Relevant documents

Epidemiologic indicators

Location
Date
Disease
Host
Number of cases
Information extraction

- Based on text mining methods
  - Rule-based approach (labelled corpus)
  - SVM classification

<table>
<thead>
<tr>
<th>Type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>0.80</td>
</tr>
<tr>
<td>Dates</td>
<td>0.88</td>
</tr>
<tr>
<td>Diseases</td>
<td>0.96</td>
</tr>
<tr>
<td>Cases</td>
<td>0.85</td>
</tr>
<tr>
<td>Species</td>
<td>0.96</td>
</tr>
</tbody>
</table>
Challenges and issues

Identification of information

Combination of information

Epidemiological knowledge

1. Data acquisition
2. Data categorization
3. Information extraction

Relevant documents

Epidemiologic indicators

Location
Date
Disease
Host
Number of cases
Identification of information: How to deal with spatial ambiguity?

- Spatial ambiguities (Amitay et al. 2004)

<table>
<thead>
<tr>
<th>Type of ambiguity</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geo/non-geo</td>
<td>‘This’, France vs ‘This’, preposition</td>
</tr>
<tr>
<td>Geo/geo</td>
<td>‘Paris’, France vs ‘Paris’, Texas</td>
</tr>
<tr>
<td>Geo/organization</td>
<td>‘the Paris-based OIE’</td>
</tr>
</tbody>
</table>
Identification of information: What are we looking for?

Symptoms
- malformation
- hyperthermia
- diarrhea
- anorexia
- weakness
- ...

Syndroms
- reproductive
- digestive
- ...

Heterogeneous data
- zootechnical data
- laboratory data
- economical
- ...

Weak signals
Combination of information

Official sources
- Source 1
- Source 2
- ...
- Source n

Unofficial sources
- Spatial entities
- Temporal entities
- Diseases
- Hosts
- Syndromes

How to combine?
Conclusion

Stagiaires: Baptiste Belot, Clément Hemeury, Max Devaud, Thomas Filol, Ibrahim Bompoko Ebongo, Rohan Goel

Autres collaborateurs: Barbara Dufour, Pascal Hendrikx, David Chavernac