



PULS

Pattern-based Understanding and Learning System

Current work

- Broaden coverage beyond infectious epidemics, to other threats to public health, including CBRN
- Improve automatic quality assessment of extracted facts by estimating confidence and relevance
- Expand the number of languages covered by the system

Mission

PULS supports Health Authorities through fully automatic analysis for tracking emerging threats to public health:

- **extracting facts** from on-line news, using language technology and semantic analysis
- **aggregating** similar stories across multiple news sources, to reduce redundancy
- automatic **quality assessment**, to rank reports according to urgency



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Department of Computer Science
University of Helsinki
Finland

SA treats 68 cholera patients on Zimbabwe border

South Africa has treated 68 cholera patients since the weekend in a town by the border with Zimbabwe, where the disease has killed dozens of people in recent weeks, a health official said today. "Since Saturday, we have received and treated a total of 68 cholera patients from Zimbabwe," said Phuti Selobi, spokesman for the health department in the town of Musina said. "Sixty-six of them are Zimbabweans while two others are South Africans engaged in cross-border business," Selobi told AFP. "Only 14 of them are still in the hospital," he added, noting that no one has died of cholera in South Africa.



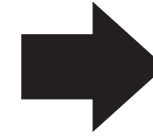
SA treats 68 cholera patients on Zimbabwe border

South Africa has treated **68 cholera patients** since the weekend in a town by the border with Zimbabwe, where the **disease has killed dozens** of people in recent weeks, a health official said today. "Since Saturday, we have received and treated a total of **68 cholera patients from Zimbabwe**," said Phuti Selobi, spokesman for the health department in the town of Musina said. "Sixty-six of them are Zimbabweans while two others are South Africans engaged in cross-border business," Selobi told AFP. "Only 14 of them are still in the hospital," he added, noting that no one has died of cholera in South Africa.



SA treats 68 cholera patients on Zimbabwe border

South Africa has treated **68 cholera patients** since the weekend in a town by the border with **Zimbabwe**, where the **disease has killed dozens** of people **in recent weeks**, a health official said today. "Since **Saturday**, we have received and treated a total of **68 cholera patients from Zimbabwe**," said Phuti Selobi, spokesman for the health department in the town of Musina said. "Sixty-six of them are Zimbabweans while two others are South Africans engaged in cross-border business," Selobi told AFP. "Only 14 of them are still in the hospital," he added, noting that no one has died of cholera in South Africa.



Disease **Cholera**
 Country **Zimbabwe**
 Location **Musina**
 Time **2008-11-20**
 Victim **People**
 Number **68**

Data Collection

News items arrive continuously in real-time, as **plain text**. **PULS** currently receives raw text from the partner system, EC-JRC's MedISys. **PULS** processes news from specialized sites, such as ProMED-Mail, and can process data from other news aggregators.

Linguistic Analysis

Plain text is analysed for indicators of epidemic-related cases. The analysis is fully automatic, employing state-of-the-art language technology and machine learning.

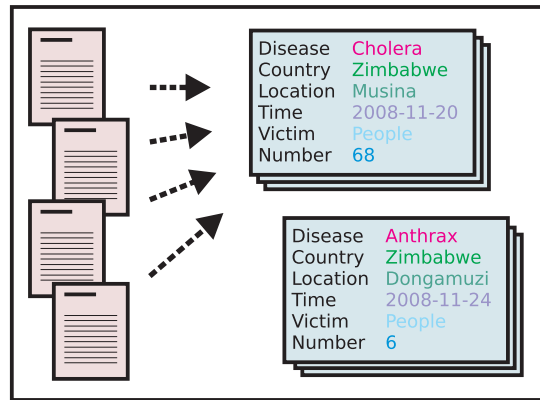
Understanding

Each detected case is transformed into **structured information**:

- **what** disease
 - **where** — country, location,
 - **when** — date
 - **who** — victims, number, ...
- and added to a database of events

Learning

PULS gives authorized users the option to provide feedback to the system, e.g., by correcting errors in the automatic analysis. The system adapts to improve the analysis on subsequent reports.



Aggregation

The same report is frequently encountered in multiple news sources, possibly with slight variations. For epidemic surveillance, the user does not want to read the same story over and over again, multiple times. **PULS** tries to reduce redundancy by:

- Gathering multiple reports on the same story, across time and multiple sources, and presenting them to the user as a **unified record**. Links to original documents are preserved
- Assigning a **relevance** score to the unified record, based on criteria that determine the urgency and importance of the event
- Linking related stories across time into one evolving event/thread

Published	Source	Disease	Country	Begin	End	Total	†	Descriptor
2008.11.14	swissinfo	Influenza	Switzerland	2008.11.07	2008.11.07	--	†	children
2008.11.10	googlenewshealth	Norovirus	USAMichigan	2008.11.06	2008.11.06	2		more than two dozen ...
2008.11.07	googlenewshealth	Norovirus	Netherlands	2008.11.06	2008.11.06	2		more than two dozen ...
2008.10.21	mednews	Tuberculosis	Turkey	2008.10.21	2008.10.21	550		550 children
2008.10.21	medicalnews	Tuberculosis	Turkey	2008.10.21	2008.10.21	550		550 children
2008.10.21	newsmedical	Tuberculosis	Turkey	2008.10.21	2008.10.21	550		550 children
2008.10.21	googlenewshealth	Hand , Foot , And Mo...	China	2008.10.20	2008.10.20	3	†	three children
2008.10.20	alertnet	Hand , Foot , And Mo...	China	2008.10.20	2008.10.20	3	†	Three children
2008.10.20	channel4news	Hand , Foot , And Mo...	China	2008.10.01	2008.10.20	--	†	the children
2008.10.20	dailypress	Hand , Foot , And Mo...	China	2008.10.01	2008.10.20	22	†	the children
2008.10.14	24dash	Tuberculosis	UK	2008.10.14	2008.10.14	7		Seven children
2008.10.20	news24	Hand , Foot , And Mo...	China	2008.10.13	2008.10.13	3	†	Three children
2008.10.20	channelnewsasia	Hand , Foot , And Mo...	China	2008.10.13	2008.10.13	3	†	Three children
2008.10.10	allafrica	Gastroenteritis	Nigeria	2008.10.10	2008.10.10	--	†	children
2008.10.07	guardian	Diarrhoeal Disease	worldwide	2008.10.08	2008.10.08	more	†	more children

1 2 3 4 5 6 ... 30 31 32 >>

Decision support

PULS stores all events in structured form in a permanent database. To support the work of the Public Health Authorities, the database provides the ability to search and query for specific diseases, countries, victims (adults/children/human/animal, ...), time period, and other attributes. The database feeds into down-stream analysis and visualization tools: maps, charts, etc.

