Global Public Health Intelligence Network (GPHIN)

Machine Translation Summit
September 2005

Michael Blench
GPHIN Technical Advisor and Project Coordinator
This presentation will discuss:

- What is GPHIN?
- How does it work?
- SARS
- Where are we going?
- GPHIN’s Machine Translation
What is GPHIN?

• A ‘solution’ comprised of a team of analysts directing a very innovative and unique technology.

• The ‘Team’:
  – Multi-ethnic
  – Multi-cultural
  – Multi-lingual
  – Multi-disciplined
  – Both genders and all age groups
Goal

• Provide to decision makers, early warning intelligence of potential public health threats to aid them in the avoidance, mitigation and management of an ‘event’.
Public Health Organizations at the International Level

- WHO and GOARN
- PHAC – Public Health Agency of Canada
- CDC – Center for Disease Control
- Others
Public Health Surveillance System

- **Regional / International level**
  - Analysis and Feedback. Support
  - Policy and targets funding

- **Central level**
  - Analyze, Investigate, Confirm, Plan, Fund, Respond, and Feedback

- **Intermediate level**
  - Analyze, Investigate, Report, Respond and Feedback

- ** Peripheral level**
  - Detect, Treat, Report
Public Health Surveillance System with GPHIN

- **Regional / International level**
  - Analysis and Feedback, Support
  - Policy, target funding

- **Central level**
  - Analyze, Investigate, Confirm, Plan, Fund, Respond, Feedback

- **Intermediate level**
  - Analyze, Investigate, Report, Respond, Feedback

- **Peripheral level**
  - Detect, Treat, Report

GPHIN – Early Warning System
Background

• GPHIN’s Role and management
  – GPHIN was developed by the Public Health Agency of Canada
  – Managed by the Agency’s Centre for Emergency Preparedness and Response

• GPHIN’s history
  – 1998/99 – Prototype developed by Health Canada in partnership with WHO
  – 2000 – Proof-of-concept
  – 2001 – Prototype deployed into production
  – 2002 – SARS
  – 2003 – Collaborative Research Agreement with Nstein Technologies Inc.
  – 2004 – Official launch of GPHIN II at United Nations
GPHIN’s Modus Operandi

Timeline / Hours – Days

Detection by GPHIN

Dissemination by GPHIN

Verification by users

Information on Public Health Risks

Intelligence on Public Health Risks

- Over 10,000 sources of information
- Arabic, English, French, Spanish, Russian, Chinese (Traditional & Simplified), Farsi
WHO Intelligence & Verification Activities
01 Jan 2004 - 30 Jun 2005
Total events - N=513; Verified - n=384
• New Features:
  – Viewing in 8 languages.
  – Meta data querying
  – Immediate email alerts.
  – Robust, secure system
GPHIN Taxonomy Topics

Human diseases – e.g. ebola
Animal diseases – e.g. streptococcus suis
Plant diseases – e.g. Asian soybean rust
Other Biologics – e.g. anthrax
Chemical incidents – e.g. chlorine plume
Radioactive exposures – e.g. Chernobyl disaster
Unsafe products – e.g. Prozac, NSAID
Natural Disasters – e.g. tsunami, 26 Dec 04
GPHIN’s Reporting suggestive of an atypical event (SARS?)
Cumulative number of countries reporting SARS as of May 15, 2003 by GPHIN and WHO

Note: China excludes Taiwan and Hong Kong
Source: GPHIN, 2003; WHO, 2003
GPHIN's Versatility – Information for Action

- **Detect** potential public health threats
- **Estimate** magnitude of the public health threat
- **Determine** geographic distribution of public health threat
- **Identify** control and preventative measures considered and implemented
- **Monitor** concerns of the general public
- **Observe** situational politics
Next Steps: Medium Range

• English <-> Portuguese translation software
• Enhanced translation engine
• Summarizer
• Enhanced Search
• User profiles
• Text–to–Speech
• Basic GIS (drill down mapping)
Next Steps: Long Range

- Enhanced GIS
- Other languages (e.g. Asian)
- Predictive modelling
- Economic intelligence for predicting public health events
- Advanced search – Boolean, meta data & phonetics
- Speech–to–Text (Speech recognition)
- Speech–to–Speech
GPHIN Workflow

Data Collection

Advanced Analytics

Multilingual Real Time Intelligence Rendering

Multilingual Sources

- Duplicate elimination
- Metadata Insertion (taxonomy)
- Categorization
- Translation
- Relevency Score

Multilingual Sources Languages:
- En
- Fr
- Fa
- Es
- Ru
- Ar
- Zhs
- Zht

0
30
85
90
100

Alert
Publish
Manual Analysis
Reject

Query
Output Rendered
Machine Translation

• Gatekeeper / Analysts with topical expertise plus linguistic skills provide human ‘quick’ translation of ‘GISTS’ to improve ‘comprehensibility’

• “Best-of-Breed” approach

• Dictionaries / lexicons
  – Corpora
  – Taxonomy
    • Root term
    • Synonyms
    • Spelling variants of root terms and synonyms

• Meta Data – trans-lingual text mining
Gist Edit

**Title**
Aggressive visitors carry alarming buzz: The bees are known for their potential to swarm, then attack.

**Text**
Aggressive visitors carry alarming buzz: The bees are known for their potential to swarm, then attack.

MANATEE
They can deliver 1,000 stings in one minute.
People who have died after being stung by them routinely have 40 stings per square inch on every bit of their exposed skin.
And now the African honeybee - the killer bee made only slightly more monstrous by Hollywood - has been discovered in the Tampa area and could be in Manatee County, said a University of Florida entomologist who has studied them in Costa Rica and in their native Africa.
"It's hard to tell if they are in Manatee County and I wouldn't venture a guess," associate professor Glenn Hall said by phone Monday. "It will probably be a year or two if they aren't here already. They are getting established."
The killer bee is actually slightly smaller than the European bee, which is used by beekeepers in Manatee County for honey and drop pollination. The African bee also has 10 percent less venom, Hall said.
But what it loses in size, it makes up in numbers.
"It is the large number of stings that cause the problem," Hall said. "We usually say 1,000 stings is a lethal dose and they can deliver that in one minute."
The African bee is exceptionally defensive. If its colony is approached, even by a passerby who means no harm, the bees have been known to pursue a fleeing subject for a quarter of a mile, Hall said.
Don't panic, Hall advises, even though his stories of the bee are alarming.
"A European bee will go up to 100 and quit," Hall said.
Hall has been able to create genetic markers for the African bee. He can extract the DNA from dead bees to analyze if they are the variety with attitude.
Hall's advice to Manatee County: "Right now I wouldn't worry about it," he said.
"We've had some stinging incidents (in Florida), but most have been with animals."
When African bees leave the port in Florida, they have no choice but to mate.
GPHIN News Feeds: unvalidated, publicly available

- Poetic licence or polysemic terms
- Deliberate / accidental mis-use of words
- Mis-spellings
- Local vocabulary
- Style of prose
- Presence / absence of diacritics
- Abbreviations
- Spacing of letters of a word
- Capitalization
- Chinese grammar rules
- Quirks & twists
Poetic licence or polysemic terms

- Example #1
  - “No beef in cattle bailout”
  - “Aucun boeuf dans le saut du bétail en parachute”

- Example #2
  - “Injured bats in rabies alert”
  - “Bates heridos en alerta de rabia” vs “Murcielago heridos en alerta de rabia”
Poetic licence or polysemic terms cont’d

• Example #3
  – … after drinking contaminated water in southwestern Turkey
  – … après l'eau contaminée de boisson dans la Dinde …

• Example #4
  – Officer takes a spill as Bush takes birthday spin
  – Prises de l'officier qu'un déversement comme l'anniversaire des prises du Buisson tourne
Deliberate / accidental mis-use of words

**Example #1**
- “China pesa demasiado” = “China weighs too much”

**Example #2**
- “Aggressive visitors carry alarming buzz …”

**Example #3**
- “The agriculture with the international …” (the ‘Ministry of Agriculture’ is implied)
Mis-spellings

• Example #1
  – “Global News Bites” = “Las Noticias Globales Muerden” vs
    “Global News Bits” = “Las Noticias Globales Pulsaciones ”

• Example #2
  – “Penguin Sex Deaths”
Local Vocabulary (slang)

• Example #1
  – “anthrax”, proper Russian term = сибирская язва
  – Local term = сибирк*

• Example #2
  – “Time for cha” vs “Time for tea”

• Example #3
  – Proper term: Tuberculosis------
    Local term: Tuberculosis------
  – Proper term: Chickenpox--------
    Local term: Chickenpox--------
Bad Example:

• ""

Translates to:

• The child sets up green grass of the night reshahar doubting that others reports on a poisoning giving vent to self's anger is hurt heavily and in large amount "innocent" state-owned forest farm large area being killed with poison
Good Example:

- 

Translates to:

- By report, Honda FCX accords with all USA government safety standards. It Driving way is same as average automobile, noise is smaller than ordinary vehicle, in the fuel cell being pretending compresses a hydrogen but being not that the gasoline, blowoff are complete is water, Jiaman hydrogen one. Can run 300 kilometres time.
Presence / absence of diacritics

• **Example #1**
  – tâche = task
  – tache = stain, speck

• **Example #2**
  – When an apostrophe replaces a letter and the next letter has a diacritic:
  – “.. à l'énergie atomique ..” = “.. at l 'atomic energy ..”
Abbreviations

• Example #1
  – РОССЕЛЬХОЗНАДЗОР ГАРАНТИРУЕТ БЕЗОПАСНОСТЬ ..
  translates to: ₸ÎÑÑÅËÜØÛÏÇÍÀÄÇÎÐ GUARANTEES SAFETY ..
  РОССЕЛЬХОЗНАДЗОР is the Russian Agriculture Control Authority

• Example #2
  – .. почти до 2 тыс (thousand) гектаров .. = .. up to 2 òûñ hectare ..
Spaces within a word

- **Example #1**
  - “С Р О Ч Н О Министры стран …” where
  - “С Р О Ч Н О” (with spaces between letters should be “U R G E N T” but translates to:
  - “WITH R ABOUT CH N ABOUT Ministers of the countries …”

- **Example #2**
  - “.. Shijiazhuang, **May 1** XinHua ..”

(13 May) =
Capitalization (in place names)

• Example #1
  – Армения (Armenia) = ԱՐՄԵՆԻЯ

• Example #2
  – Ереван (Yerevan, capital of Armenia) = ԵՐԵՎԱՆ
Chinese grammar rules

Gist Edit

Title
Aggressive visitors carry alarming buzz; The bees are known for their potential to swarm, then attack

Text
Aggressive visitors carry alarming buzz; The bees are known for their potential to swarm, then attack

MANATES
They can deliver 1,000 stings in one minute. People who have been stung by them routinely have 40 stings per square inch on every bit of their exposed skin.

And now the African honey bee - the killer bee made only slightly more monstrous by Hollywood - has been discovered in the Tampa area and could be in Manatee County, said a University of Florida entomologist who has studied them in Costa Rica and in their native Africa.

"It's hard to tell if they are in Manatee County and I wouldn't venture a guess," associate professor Glenn Hall said by phone Monday. "It will probably be a year or two if they aren't here already. They are getting established.

The killer bee is actually slightly smaller than the European bee, which is used by beekeepers in Manatee County for honey and crop pollination. The African bee has a 10 percent less venom, Hall said.

But what it loses in size, it makes up in numbers.

"It is the large number of stings that say 1,000 stings is a lethal dose and The African bee is exceptionally dedicated passerby who means no harm, they are subject for a quarter of a mile," Hall said.

Don't panic, Hall advised, even though his stories of the bee are alarming.

"A European bee will pursue 60 to 100 feet and quit," Hall said.

Hall has been able to create genetic markers for the African bee. He can extract the DNA from dead bees to analyze if they are the variety with attitude.

Hall's advice to Manatee County: "Right now I wouldn't worry about it," he said.

"We've had some stinging incidents in Florida, but most have been with animals."

When African bees leave the ports in Florida, they have no choice but to mate.
Quirks and Twists

• Example #1
  – AIDS: ---- ------ “love growing disease”

• Example #2
  – تيديدة Hodeida: A city in Yemen
  – تيديدة Hadida: Iron in Arabic
  The reader has to understand the sense of the phrase to decide how to pronounce the word

• Example #3
  – يييباصملإ ف اعسا ييف فمأو قل و هجملإ يم جلإب دقف و اتلاح translates to:
  Death cases by the unknown fever .. When it should translate to TWO cases
Integration Issues

- Awkward API’s
- Poor documentation
- Lack of standards for dictionaries / lexicons
- Bugs
- Clashes between different manufacturers
Operational Issues

- Inherently unstable code
- Bugs
- Crashes
- Unpredictable performance
nTranslator™

• Overcomes some of the integration and operational issues:
  – Normalizes API’s
  – Detects ‘engine’ crashes and re-boots
  – Overcomes in-compatibilities
  – Solves display / rendering problems
  – Converts file formats
Recommendations

- Translation memory
- Transliteration
- Tighter code to improve speed and reduce cycles
- Standardization
Contact Information

• Ms. Abla Mawudeku, Manager, GPHIN
  (613) 957-2715
  Abla_Mawudeku@phac-aspc.gc.ca

• Michael Blench, GPHIN Technical Advisor
  (613) 957-3567
  Michael_Blench@phac-aspc.gc.ca

• Ms Isabelle Desjardins, GPHIN Account Manager
  Nstein Technologies Inc.
  (514) 908-5406
  isabelle.desjardins@nstein.com