



Machine Assisted Medical Device Signal Detection Framework using Text Mining Techniques

Sithu D Sudarsan^{1,2}, Raoul P Jetley¹,
Brian Fitzgerald¹ and Srinivasan
Ramaswamy²

¹ Division of Electrical and Software Engineering,
Office of Science and Engineering Laboratories,
Food and Drug Administration,
Silver Spring MD, USA

² University of Arkansas at Little Rock
Little Rock AR, USA



Outline

- Introduction
- Signal Detection Framework
- Implementation
- Experimental results
- Future work
- Discussion

- Introduction
- Signal Detection Framework
- Implementation
- Experimental results
- Future Work
- Discussion

The Organization

- Food and Drug Administration (FDA)
 - US regulatory agency for food and drugs
 - Food, Drugs, Devices, Biologics, Veterinary, Cosmetics and Tobacco
 - Ensure safety and effectiveness
 - Premarket approval and post market surveillance
 - Managed through centers and field offices
 - CDRH¹ – Center for Devices and Radiological Health
- ¹ <http://www.fda.gov/MedicalDevices/default.htm>

Mining for *signals!*

“Refers to the detection of relationship between a drug (or medical device) and its adverse effects either previously unknown or a change in the nature of occurrence of such relationship” – WHO



Where to mine?

CDRH Document repository

Data	Discovery	Extraction
Structured	Data Mining	Database Query
Unstructured	Semantic Mining	Extraction / Retrieval



ISDS 2009 Dec 3, 2009

- A term with specific meaning in the context of health monitoring and surveillance.
- “Refers to the detection of relationship between a drug and its adverse effects either previously unknown or a change in the nature of occurrence of such relationship” – World Health Organization

ISDS 2009 Dec 3, 2009



CDRH document repository*



- Premarket submissions (manufacturers)
- Adverse event reports (multiple sources)
- Annual reports (manufacturers)
- Recalls (FDA issued)
- Listing and Registration (FDA compiled)
- Approved standards and guidance documents (multiple sources)
- Establishment Inspection Reports (internal)
- Inspectional Observations (internal)

* Only documents relevant to the discussion are listed

ISDS 2009 Dec 3, 2009



Characteristics



- Medical Device adverse event Reports (MDR) account for > 100,000 documents each year
- Accrued over several decades and keeps growing
- Hard copies of documents are scanned and stored (mostly *pdf* files; also plain text and rich media formats)
- 1 to 100+ pages, KB to several MB

ISDS 2009 Dec 3, 2009

- Introduction
- **Signal Detection Framework**
- Implementation
- Experimental results
- Future work
- Discussion

Signal detection¹ in devices

- Done by analyzing received MDRs, pre-market review observations, literature reports/presentations, failed inspections, in-house research, etc.
- Types of adverse events
 - Related to – patient, operational context
 - Use trend, statistics, proportional reporting ratio
 - Related to – device design, manufacturing, ...
 - Review design documents

¹<http://www.fda.gov/Safety/FDAsSentinelInitiative/default.htm>

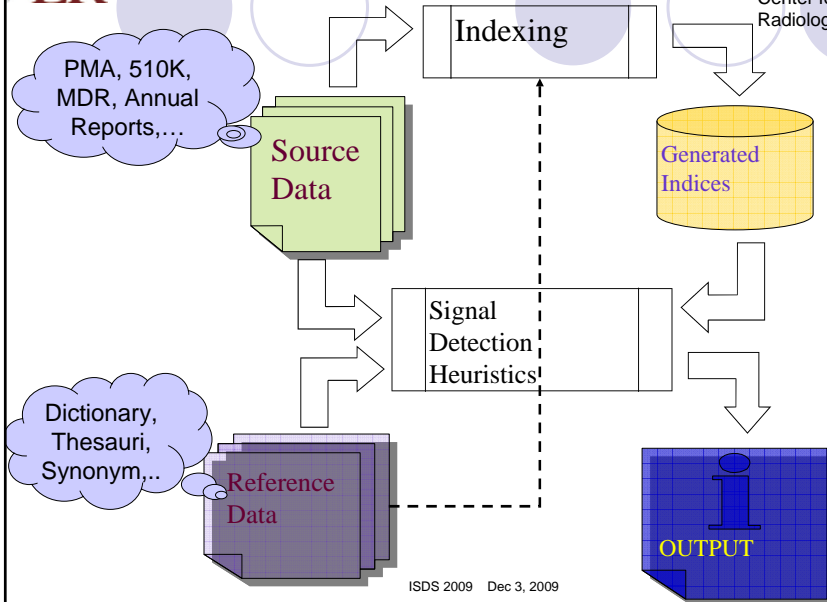
Early, automated

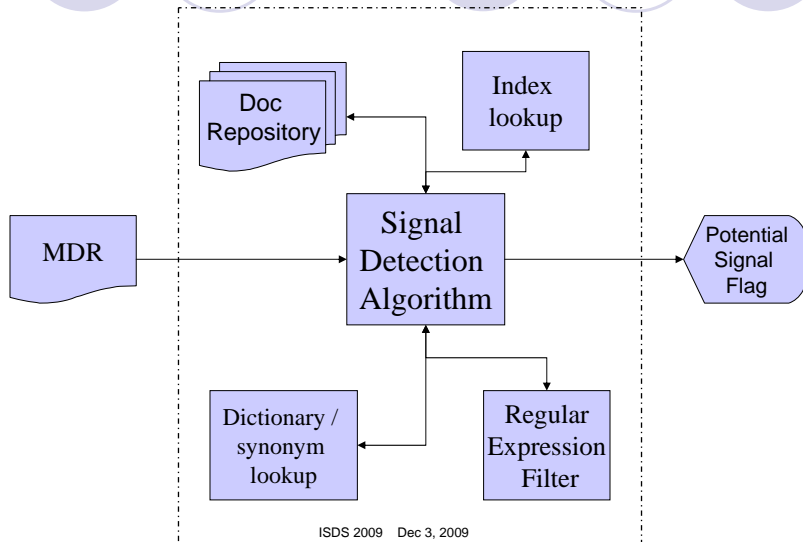
- Signal prediction support
 - Use text analysis to predict potential signals
- Signal evaluation support
 - Helps reviewer to confirm potential signal by advanced search / query operations
 - Retrieves relevant documents

- Repository related
 - Handle huge growing document collection
 - Over 250,000 added each year (CDRH only)
 - Support dictionaries / thesauri / synonyms
 - Not depend on Annotation & Ontology
 - Handle poor quality OCR'd documents
- User related
 - Diverse user group
- Framework related
 - Distributed data
 - Scalability and fault-tolerance

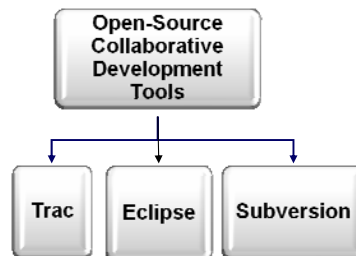
- Introduction
- Document repository
- Framework considerations
- **Implementation**
- Experimental results
- Future Work
- Discussion

IR framework overview





- Lucene (lucene.apache.org)
- PDFbox (incubator.apache.org/pdfbox)
- Tomcat (tomcat.apache.org)
- POI (poi.apache.org)
- UIMA (incubator.apache.org/uima)
- JDK (java.sun.com)

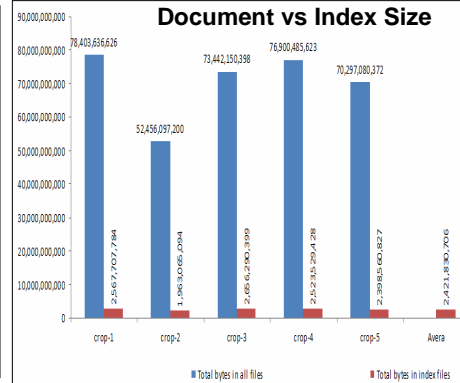
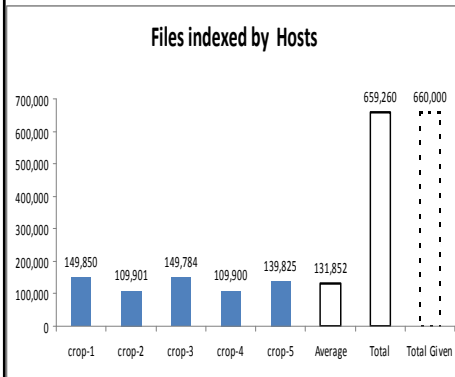


- Introduction
- Signal Detection Framework
- Implementation
- **Experimental results**
- Future work
- Discussion

Observations

- Indexing with a dual-quad core machine, 32-bit OS (3 GHz)
 - Up to 100,000 documents
 - 1,000 documents / min; 400 MB/min
- Indexing in grid (parallel) environment – 5 dual-quad core machines, 64-bit OS (3 GHz)
 - Up to 3,250,000 documents
 - 2,500 documents / min / machine; 1GB/min

Files to be indexed: 660,000
 Number of machines used: 5
 Analyzer: whitespaceanalyzer
 Stopword removal: no



- Introduction
- Document repository
- Framework considerations
- Implementation
- Experimental results
- Future Work
- Discussion



Future Work



Ongoing implementations

- Multi-searcher
- Custom scoring
- Dictionary support

Planned implementations

- Proactive signal detection heuristics
- User defined scoring
- Multi-repository support

ISDS 2009 Dec 3, 2009



- Introduction
- Document repository
- Framework considerations
- Implementation
- Experimental results
- Future work
- **Discussion**

ISDS 2009 Dec 3, 2009



Acknowledgements



- The authors would like to acknowledge the support provided by colleagues at
 - CDRH, FDA
 - UALR
 - java-user@lucene.apache.org

ISDS 2009 Dec 3, 2009



Thank you!



- Questions, if any...

- Contact:
 - sithu.sudarsan@fda.hhs.org
 - raoul.Jetley@fda.hhs.org
 - brian.fitzgerald@fda.hhs.org
 - srini@ualr.edu

ISDS 2009 Dec 3, 2009