



---

# Recurring Anomaly Detection System – Text Clustering & Recurring Anomaly Identification

Dawn McIntosh  
Intelligent Systems Division  
NASA Ames Research Center

Nov. 9, 2006



# Problem Introduction



NASA programs have large numbers (and types) of problem reports.

- Typical Datasets: 3000+ to 27000+ records, 1-4 pages each;
- Aviation Safety Reporting System (ASRS): 27000+ records, 1 paragraph each.

These free text reports are written by a number of different people, thus the emphasis and wording vary considerably

With so much data to sift through, analysts (subject experts) need help (a) identifying possible safety issues or concerns, and (b) confirming that important problems haven't been overlooked.

- Unsupervised clustering is the initial step to accomplish this;
- We think we can go much farther, specifically, identify possible **recurring anomalies**.
  - Recurring anomalies may be indicators of larger systemic problems.



# Searching for Recurring Anomalies



Enabling the discovery of anomalous trends in complex aerospace systems

Mishap Anomaly - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

https://nx.aen.nasa.gov/nxpub/dsweb/Get/Document-35264/InterfaceTest.html

Databank

- ASRS
- CARSText\_trial
- MSFC
- PRACA\_ISS
- SCR
- SOPDSS\_NT
- SOPDSS\_T
- SPN

Context ALL Content Hits/Databank 5 QUERY

Text Mining Status Text Mining Results Text Mining Algorithms:  Classification  Clustering & ReADS TEXT MINING

**Text Mining System**  
Ames Research Center

Advanced Exploration Network Copyright © 2006

Done nx.aen.nasa.gov

**ReADS team:**

Ashok N. Srivastava, Ph.D.  
Dawn M. McIntosh  
Eugene Turkov  
Francesca Barrientos, Ph.D.

**NX team:**

David Bell, Ph.D.  
Michael Crawford

**Research sponsored by:**

NASA Engineering and Safety Center



# Searching for Recurring Anomalies

---



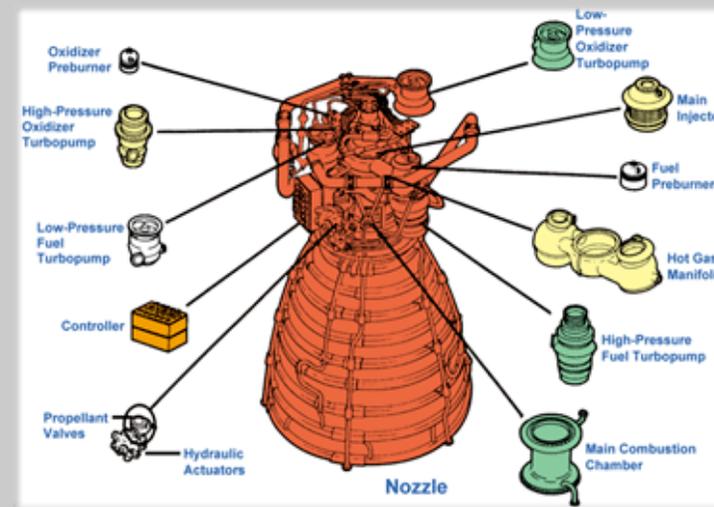
- These reports *do not* have an anomaly category associated with them.
- Potentially several hundred thousand reports.
- Some systems have been around for decades.
- Enables analysis of trends of anomalies (trending).
- Can't be addressed using standard clustering techniques.
- Our systems use content-based similarity as well as statistical similarity.



# Definition of Recurring Anomalies



- Recurrent failures described in text reports.
- Problems that cross traditional system boundaries.
- Problems that have been accepted by repeated waivers.
- Discrepant conditions repeatedly accepted by routine analysis.
- Events with unknown causes.





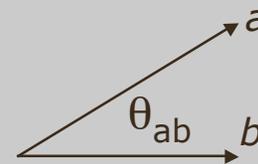
# Detecting Recurring Anomalies



1. Calculate cosine similarity between all document vectors.

	Term 1	Term 2	Term 3	Term 4	...
Doc a	3	2	1	5	...
Doc b	0	1	4	1	...
...	...	...	...	...	...

$$\cos \theta_{ab} = \frac{\langle a, b \rangle}{\|a\| \cdot \|b\|}$$

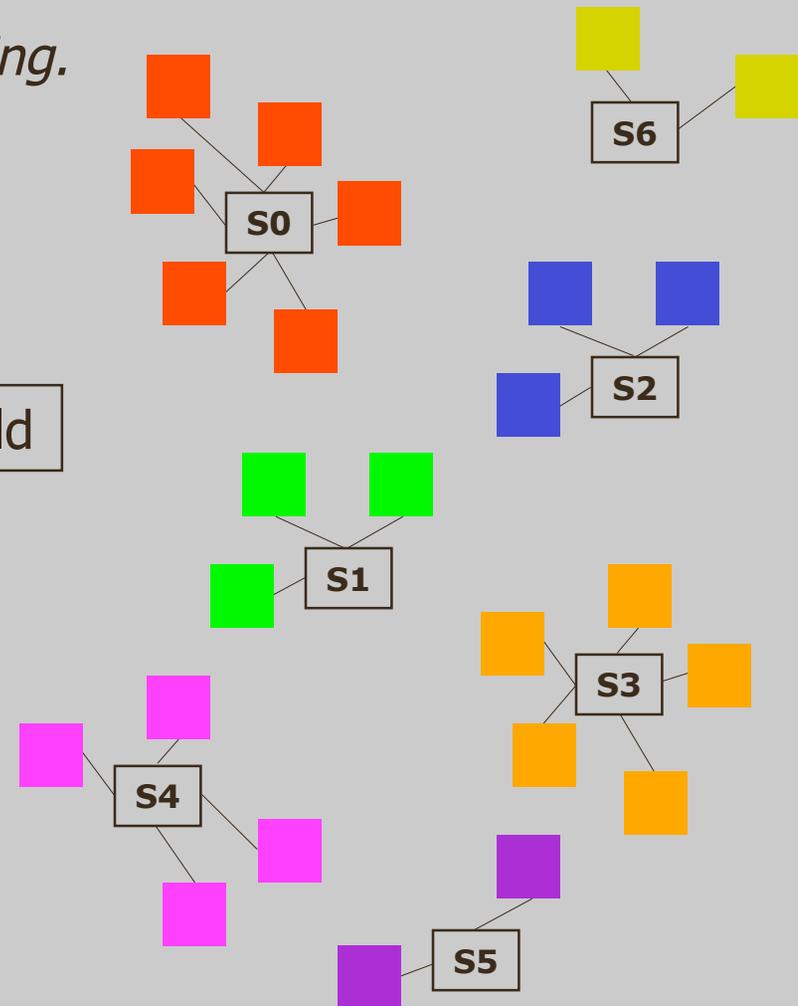
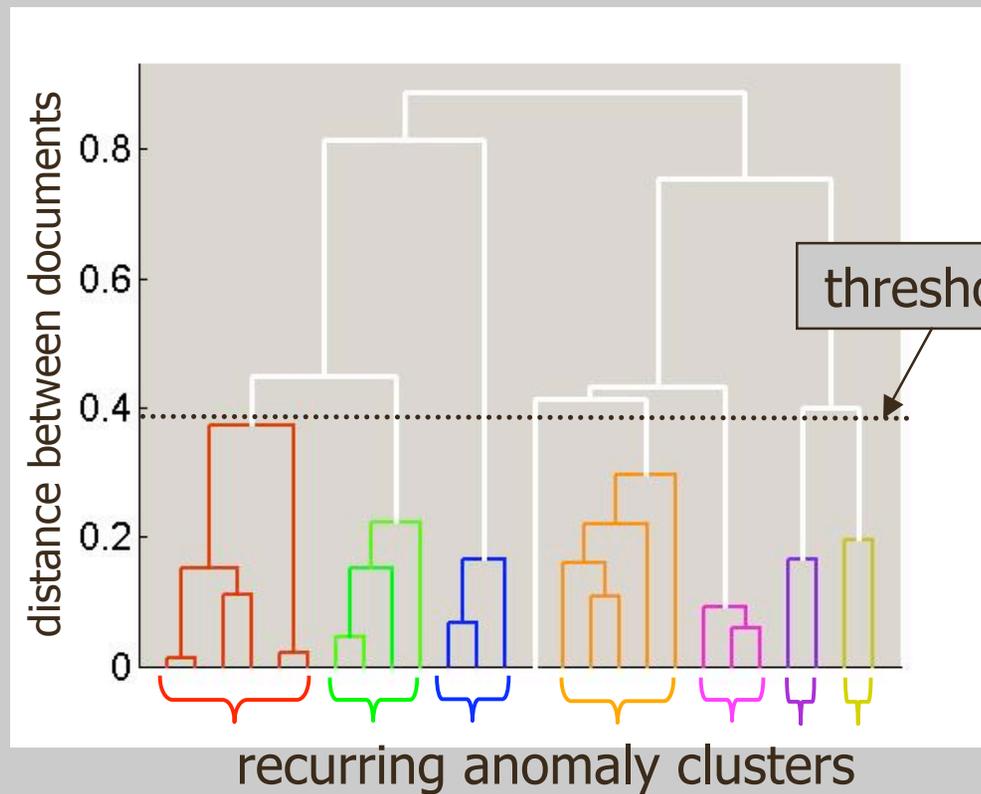




# Detecting Recurring Anomalies



2. Apply agglomerative clustering.

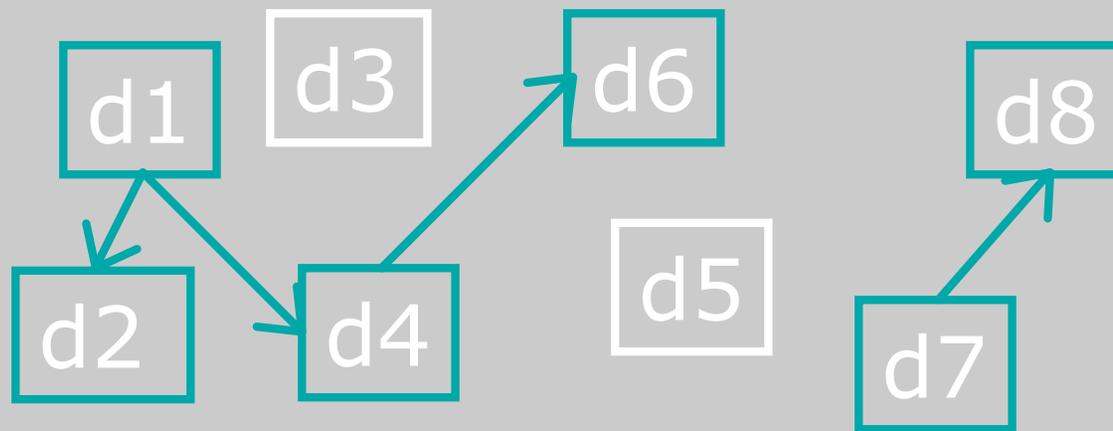




# Detecting Recurring Anomalies



*3. Identify referenced documents.*



If d1 refers to d2 and d4, and d4 refers to d6, then d1, d2, d4, & d6 are considered a recurring anomaly.

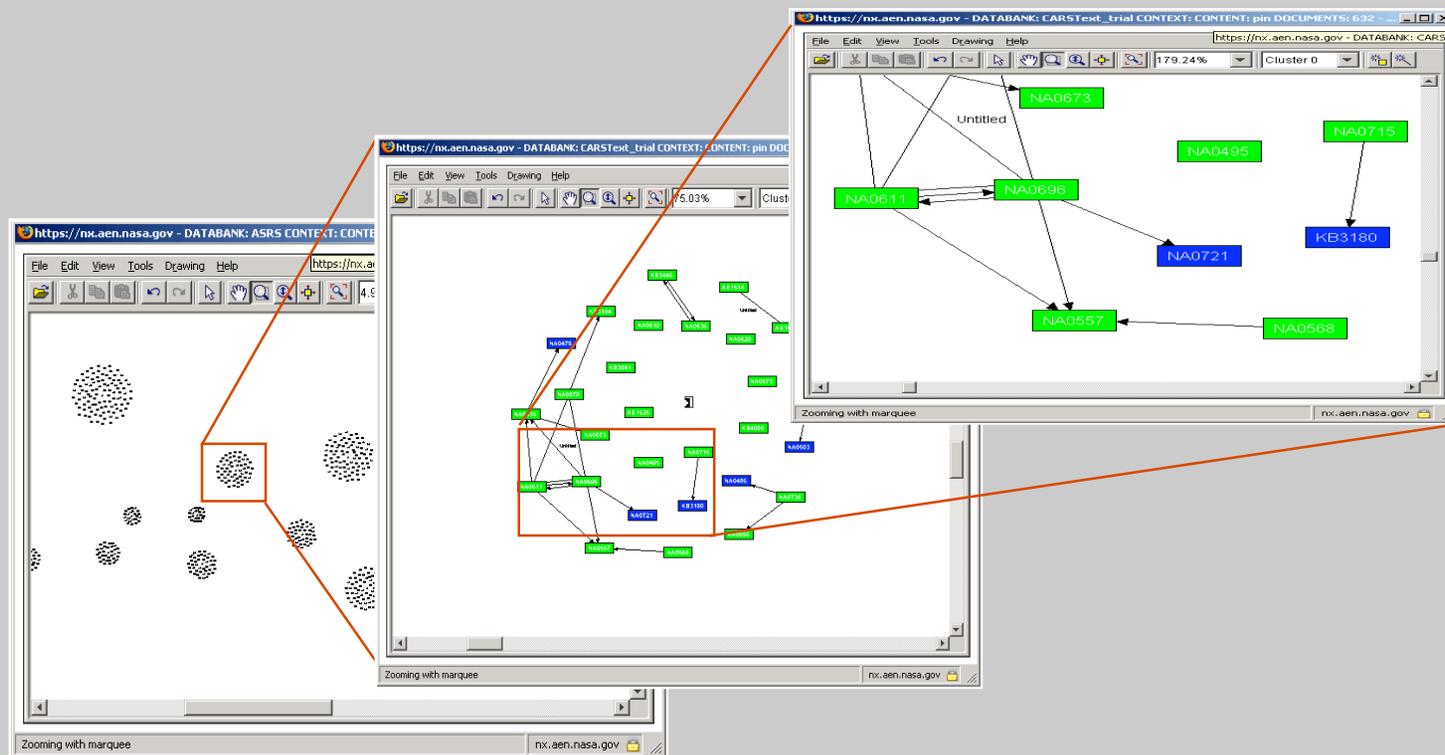
And if d7 refers to d8, then d7 & d8 are considered a second recurring anomaly.



# Detecting Recurring Anomalies



## 4. Identify & visualize possible recurring anomalies.





# Testing the Recurring Anomaly Detection System (ReADS)

---



- Subject experts reviewed a subset of the Shuttle Orbiter Corrective Action Records (CARs) specifically to identify recurring anomalies.
- We extracted 333 reports to test the performance of our system - Recurring Anomaly Detection System (ReADS).
- Of those 333 reports, the experts identified 20 recurring anomalies and ReADS identified 39 recurring anomalies.



# Types of Cluster Matches & Mismatches



## Exact Cluster Match



## Combined Clusters



## Separated Clusters





# Performance of ReADS



On a subset (333) of the Shuttle Orbiter Records:

58% of the records were eliminated as non-recurring anomalies by ReADS.

12 exact matches between recurring anomalies (RAs) discovered by experts and RAs discovered by ReADS.

6 previously unidentified RAs discovered by ReADS which were confirmed by experts.

1 record was identified by experts as being part of an RA and was missed by ReADS.

5% of the expert RAs were separated by ReADS into more than one RA.

8% of the ReADS RAs combined two expert RAs into a single RA.



# Our Innovations

---



- Enable analysis of anomaly trends using a combination of content and statistical search methods.
- ReADS is a novel tool designed especially for identifying recurring anomalies across multiple databases.
- Development of robust platform to analyze and visualize recurring anomalies.



# NX/ReADS Demonstration



Mishap Anomaly - Mozilla Firefox  
File Edit View Go Bookmarks Tools Help  
https://nx.aen.nasa.gov/nxpub/dsweb/Get/Document-35264/InterfaceTest.html

Databank  
 ASRS  
 CARSText\_trial  
 MSFC  
 PRACA\_ISS  
 SCR  
 SOPDSS\_NT  
 SOPDSS\_T  
 SPN

Context ALL Content Hits/Databank 5 QUERY

Text Mining Status Text Mining Results Text Mining Algorithms:  Classification  Clustering & ReADS TEXT MINING

**Text Mining System**  
Ames Research Center

Results may not be complete if ALL is not selected for Hits/Databank.

Databank(s): ASRS/  
ASRS/  
CONTENT: ice and storm

119416.xml  
Description  
ENRTE SFO-LAS WE ENCOUNTERED A LARGE AREA OR TSTMS ALONG OUR FILED ROUTE OF J-92. WE ADVISED ZLA THAT DEVIATIONS WOULD BE NECESSARY. WE REQUESTED AND HAD APPROVED AN EASTERLY HDG WHICH WOULD KEEP US N OF THE WX. CENTER APPEARED TO BE CONCERNED THAT OUR REQUIRED DEVIATIONS MIGHT EVENTUALLY CAUSE A CONFLICT WITH EDWARDS RESTR AIRSPACE. CENTER ADMITTED THAT THEY DID NOT HAVE A CLEAR PICTURE OF EXACTLY WHERE THE WX WAS LOCATED NOR ITS POSITION RELATIVE TO THE RESTR AIRSPACE. WHILE WE CONTINUED TO DEVIATE TO REMAIN CLEAR OF WX, WE TOLD CENTER SEVERAL TIMES THAT WE COULD NOT TURN RIGHT TOWARDS J-92 AND WHAT WERE THEIR PLANS FOR US. CENTER'S ONLY CONCERN SEEMED TO BE TO KEEP US AWAY FROM RESTR AIRSPACE. CENTER KEPT ASKING US WHEN WE COULD TURN RIGHT TOWARDS J-92. WE TOLD THEM IT WOULD BE AT LEAST 15 MI BEFORE WE COULD RIGHT. CENTER THEN GAVE US A LEFT TURN HDG 330 DEG TO AVOID RESTR AIRSPACE. ANOTHER LEFT TURN FOLLOWED SO THAT WE WERE HDG SW. WE NOW HAD WX ON BOTH SIDES OF US AND WERE CLRED TO DEVIATE AS NECESSARY. WE ENCOUNTERED LIGHT TURB AND MOD ICING AND OBSERVED A LARGE OVERHANG AHEAD W/MAMMA. WE DESCENDED TO FL190 AND CONTINUED TO DEVIATE TO REMAIN CLEAR OF CELLS. I WAS BUSY OPERATING ENGINE AND WING ANTI-ICE, TRYING UNSUCCESSFULLY TO CONTACT COMPANY DISPATCH THROUGH SFO COMMERCIAL RADIO (COMMERCIAL RADIO WOULD NOT ANSWER) AND MONITORING THE RADAR. CENTER TOLD US WE HAD TO TURN RIGHT TO AVOID RESTR AIRSPACE. THE CAPT ADVISED CENTER THAT WE COULD NOT TURN RIGHT BECAUSE OF A LARGE CELL. CENTER SAID WE COULD NOT ENTER THE RESTR AIRSPACE AND HAD TO TURN RIGHT. THE CAPT DECLARED AN EMER. CENTER TOLD US TO TURN RIGHT. THE CAPT AGAIN DECLARED AN EMER AND WAS FINALLY ACKNOWLEDGED BY CENTER AND CLRED TO DEVIATE AS NECESSARY. WE WERE THEN TOLD BY CENTER TO TURN 'HARD' RIGHT BECAUSE THERE WAS 'LIVE ROUNDS' AHEAD. WE COMPLIED, DEVIATED FOR 30 MORE MINUTES AND FINALLY LANDED IN LAS. WHY CAN'T CENTER SEE THE WX? WHY COULDN'T CENTER COORDINATE WITH EDWARDS? WE WERE IN TROUBLE AND THEY WERE NO HELP! THIS WAS THE WORST **STORM** REPORTED IN THE OWENS VALLEY IN THE LAST DECADE.

132354.xml  
Description  
AFTER TKOF WE HAD UNSAFE LIGHTS ON THE LNDG GEAR. THE NOSE GEAR DID NOT COME UP (THIS WAS THE PROB). WE CYCLED THE GEAR AND AGAIN THE NOSE GEAR WAS DOWN. WE RETURNED AND LANDED AT JAX AND FOUND THE NOSE GEAR PIN WAS PULLED, MAKING IT IMPOSSIBLE TO RAISE THE NOSE GEAR. THERE HAD BEEN A FREAK **ICE STORM** IN JAX, AND THERE WAS A LOT OF **ICE** ON THE ACFT. BECAUSE OF THIS, I DID AN EXTRA GOOD PREFLT WALK-AROUND. ALL THE GEAR PINS WERE REMOVED WHEN I DID THE WALK-AROUND! WHEN I GOT BACK IN THE ACFT, THERE WAS A MECH USING THE RADIO TO TALK TO OPS ABOUT TOWING THE ACFT TO ANOTHER GATE. THE MECHS OUTSIDE THE ACFT HOOKED UP THE TOW BAR, GETTING READY TO MOVE THE AFT. THEY DID NOT MOVE THE ACFT. IT IS STANDARD FOR THE MECHS TO PULL THE NOSE GEAR PIN WHENEVER THEY HOOK UP THE TOW BAR. I FEEL THEY PULLED THE PIN WHEN THEY HOOKED UP AND DID NOT PUT IT BACK WHEN THEY UNLID THE TOW BAR AGAIN. AGAIN, THIS WAS DONE AFTER I HAD DONE MY WALK-AROUND. THERE SHOULD BE A CAUTION LIGHT FOR THIS NOSE GEAR PIN IN THE COCKPIT