# Overview

- Objective
  - Develop a framework for measuring search efficacy for the NASA Engineering Network (NEN).

- Identify concepts
  - Operational definitions, metrics/methodology

- Framework characteristics
  - Quick deployment, fast evaluation
  - Measure any change (i.e., vendor tool upgrades, new search utilities, taxonomy updates, UI improvements).

- Baseline
  - Establish initial performance of the search facility
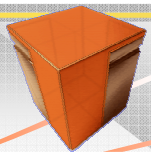 - Test the methodology

Scope

- NASA Engineering Network (NEN) search engine.
  - Research portal for NASA engineers
  - 25 repositories
  - 300,000 documents
  - 46,000 engineers
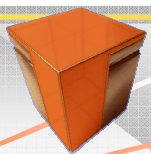 - 25 distinct engineering disciplines
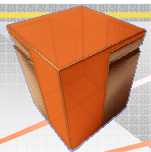
Data Generation

- Surveys
- Observations (quantitative data)
- Logs (qualitative data)

Surveys

- Task:
  - Gather background information about the user community in order to support subsequent observations.

- Software: *SurveyMonkey.com*

- Participants: *29 engineers*

- Time: *1 week*

Observations

- Tasks:
  - Conducted retrieval exercises
  - Comments/recommendations were collected
  - Perceptions and feelings were recorded
- Software: *WebEx*
- Participants: *5*
- Time: *2 weeks*
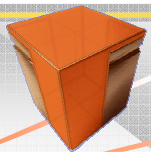
Survey

- Background
  - What are your job competencies?
  - What is your role in your current job?

- Typical Searches
  - What are the five most common topics you search for? (Unknown-item search)
  - What are the most common documents you retrieve? (Known-item search)

- Usage
  - Do you use NEN to search for documents? If not, why?
 - If you use NEN, how many times per month?
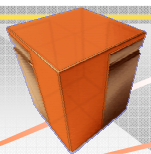
Observations

- ## Laboratories
  - JPL, Kennedy, Goddard, Langley, Ames

- ## Questions (search neutral)
  - Current projects & their associated tasks?
  - Recent problems with online search?
  - Kind of info. regularly searched? (unknown-item)
  - Specific docs regularly retrieved? (known-item)
  - How is this information typically found? Why?

- ## Activities (NEN/Verity)
  - Several unknown- and known-item searches

Statistics

- ## Known-Item search

  - Query modifications

  - Click count

  - Satisfaction score (5-point scale)

  - Average Ranked Position (ARP)

- ## Unknown-Item search

  - Relevant hits (precision)

  - Relevancy score for each item on the results page (3-point scale)
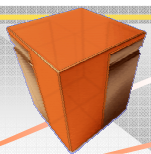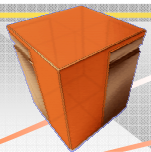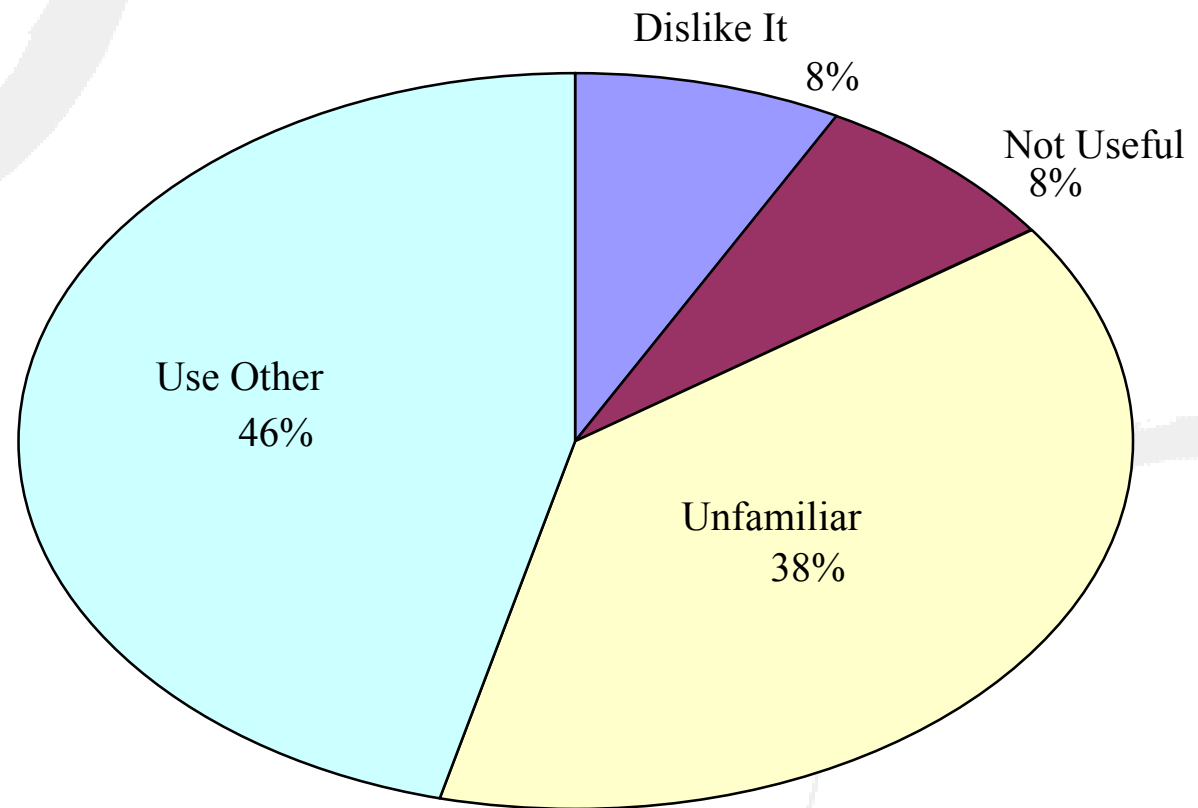
Log File Analysis

- What: *Verity search logs*

- When: *Previous 6 months*

- Types
  - clicks
  - queries

- Log views (grouped by)
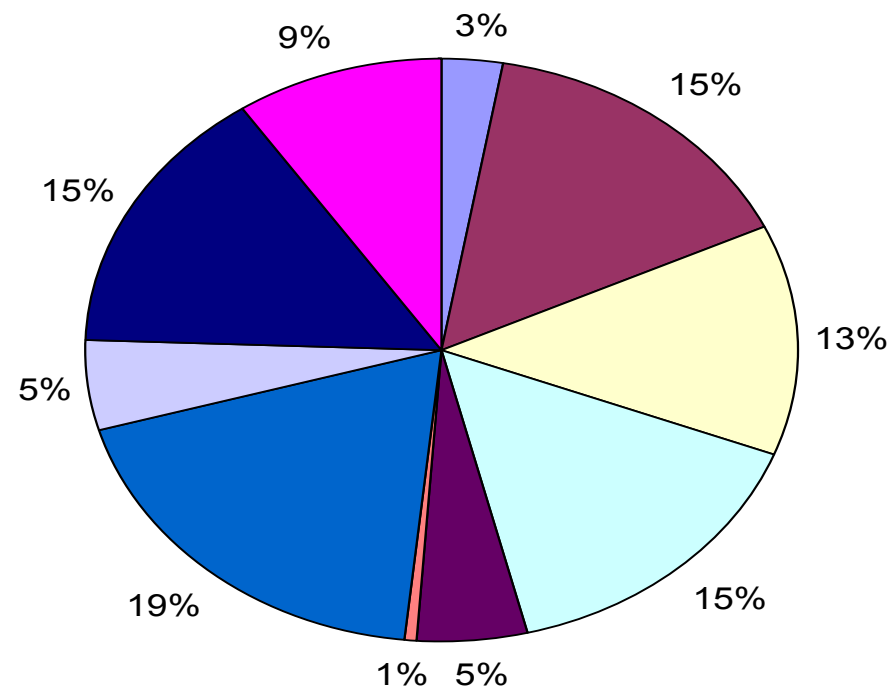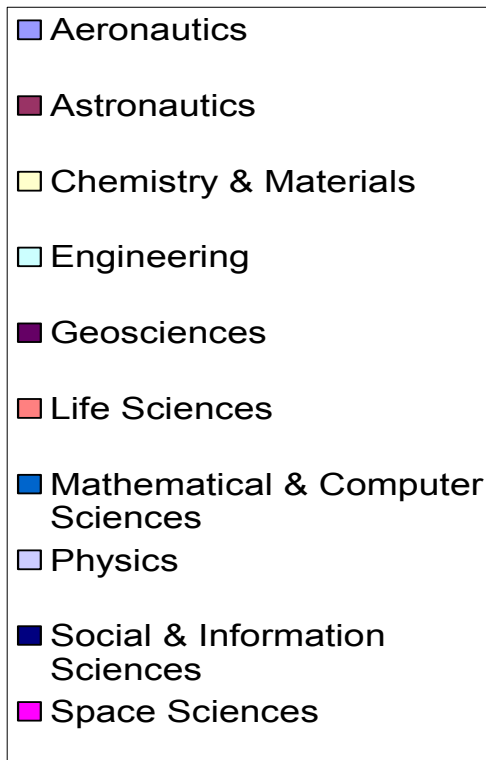 - collections
  - categories (topics)

Survey Results

Reasons for not using NEN:

Pie chart showing reasons for not using NEN: Dislike It 8%, Not Useful 8%, Unfamiliar 38%, Use Other 46%

Survey Results

Most common topics:

Aeronautics	
Astronautics	
Chemistry & Materials	
Engineering	
Geosciences	
Life Sciences	
Mathematical & Computer Sciences	
Physics	
Social & Information Sciences	
Space Sciences	

Observation Results

Metrics	Known-Item			Unknown-Item		
	Mean	Median	Mode	Mean	Median	Mode
Query Mods	0.8	1	1	0.6	1	1
Click Count	4.4	3	2	1.89	2	2
Relevant Hits	N/A	3	3	2.89	2	1
ARP	0.85	0.8	N/A (nulls)	0.31	0.3	N/A (nulls)
Relev. Score (RS)	N/A	N/A	N/A (nulls)	2	2	N/A (nulls)
Satisfaction Score	4.6	5	5	3.6	4	4

Satisfaction Scale
0. Very Satisfied – *would use it regularly*
1. Somewhat Satisfied – *would use it periodically*
2. Adequate – *may use it sometimes*
3. Somewhat Unsatisfied - *probably would not use it*
4. Not Satisfied – *would never use it*

Relevancy Scale
0. Relevant
1. Somewhat Relevant
2. Irrelevant
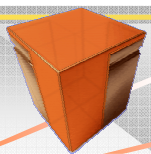
Observation Results

*Context is a problem, users feel disoriented*

Formatting
- "Results don't indicate where query terms are found in the documents."

Weighting
- "Too many results don't match."

- "The resources are not being found, only other documents referencing them."

- "There are too many different types of results."
  - Searcher can't figure out how the engine parses their queries

Indexing
- "Where are the titles coming from?"
  - "NASA Software Quality Assurance Audits Guidebook" is indexed with a title of "SMAP-GB-A301"
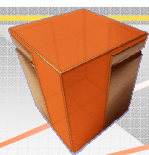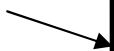
Known-Item Search

"NASA Software Assurance Standard"

NASA_STD_87398	Null
NASA 8739	No
STD 8739	No (not even STDs)
8739	No
87398	1/10
NASA-STD-8739.8	9/10

Strange Syntax

Need stemming rules based on document types.

Known-Item Search

"NASA Software Safety Standard"

	NEN	Google
7150.2	No	Top Hit
71502	No	No

Inconsistent syntax

Links to NODIS
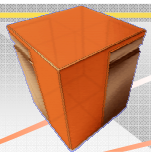
Known-Item Search
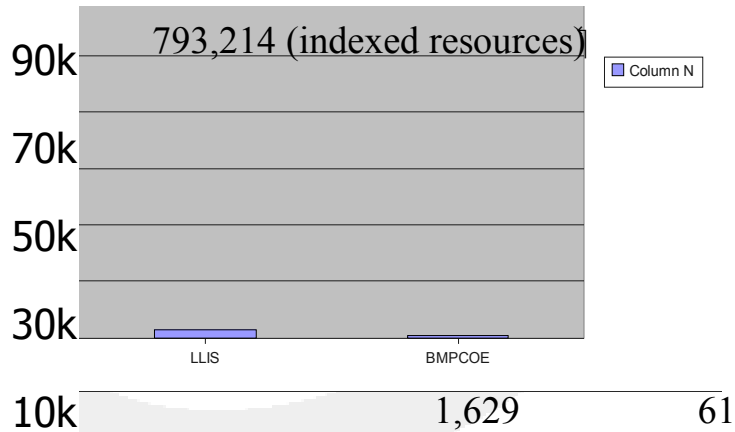
"NASA System Engineering Handbook"

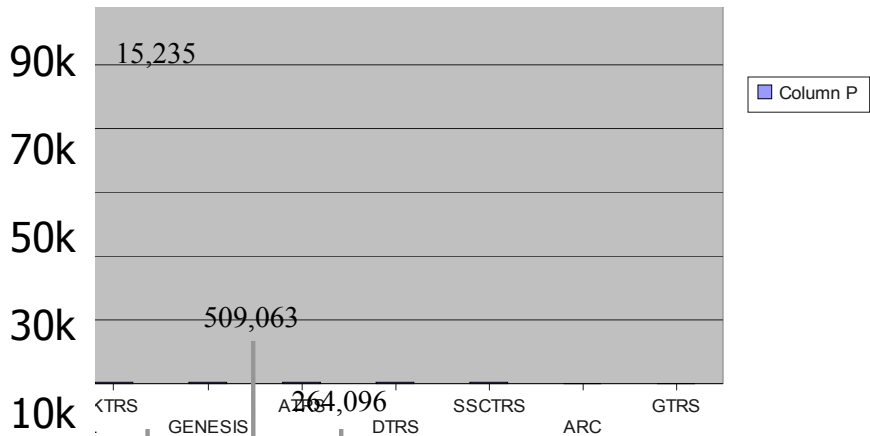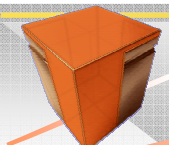| System Engineering Handbook	Top return was a 1992 draft. Official 1995 published paper - not returned
SP-6105	No (many items referencing it were found, though)

frustration

Clicks on Collections

793,214 (indexed resources)

□ Column N

90k	
70k	
50k	
30k	LLIS BMPCOE
10k	1,629 61

STI   LLIS   BMPCOE

Scientific & Technical Information - Reports

15,235

□ Column P

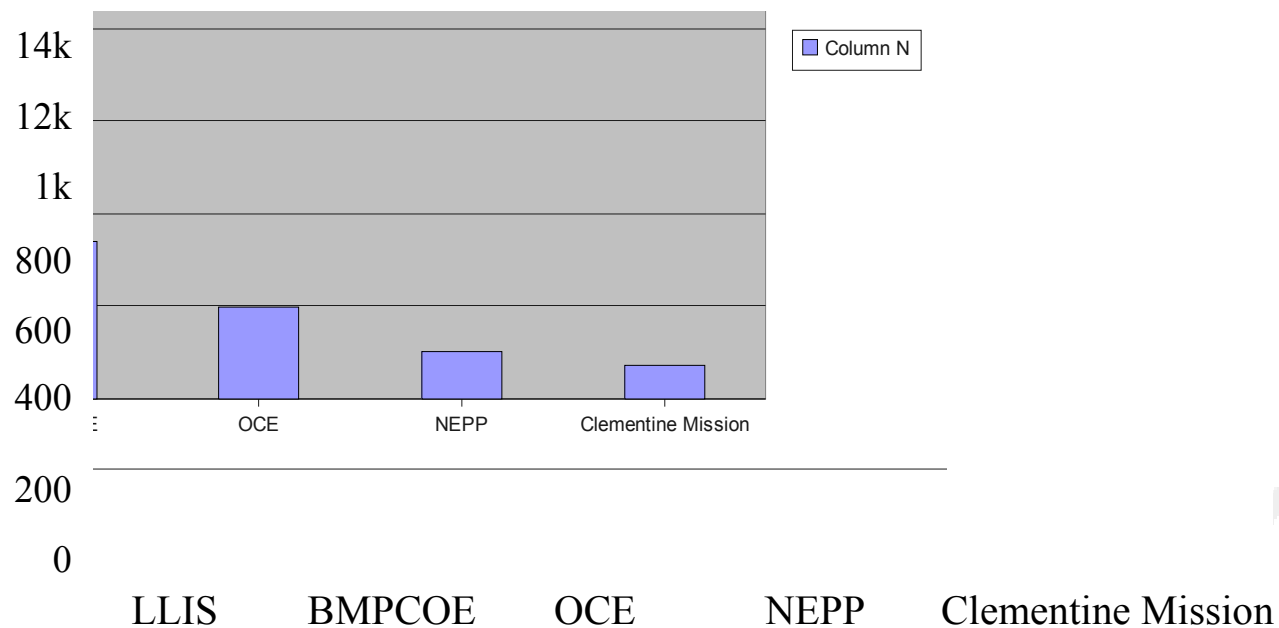| 90k | |
70k	
50k	
30k	509,063
10k	XTRS GENESIS 264,096 DTRS SSCTRS ARC GTRS

NACA

NTRS

BIOMEDCENTRAL

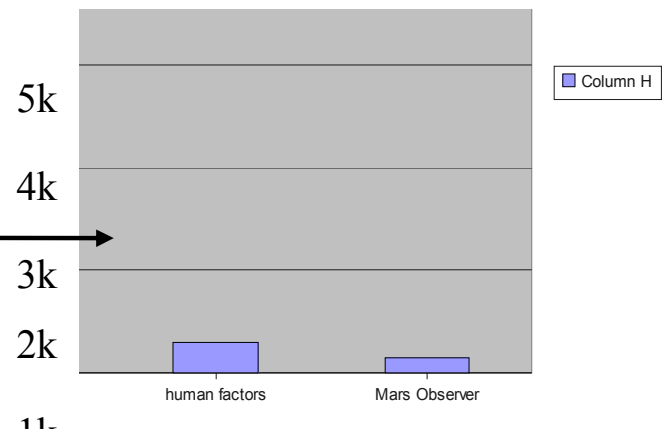# Clicks on collections (after STI)

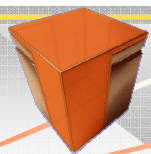*NOTE: Lessons Learned was the most popular collection during observations*
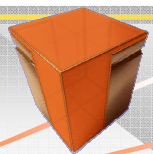
Queries in LLIS

Most document searches are failing!

- Standards
- Procedures
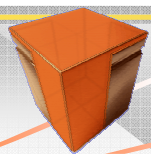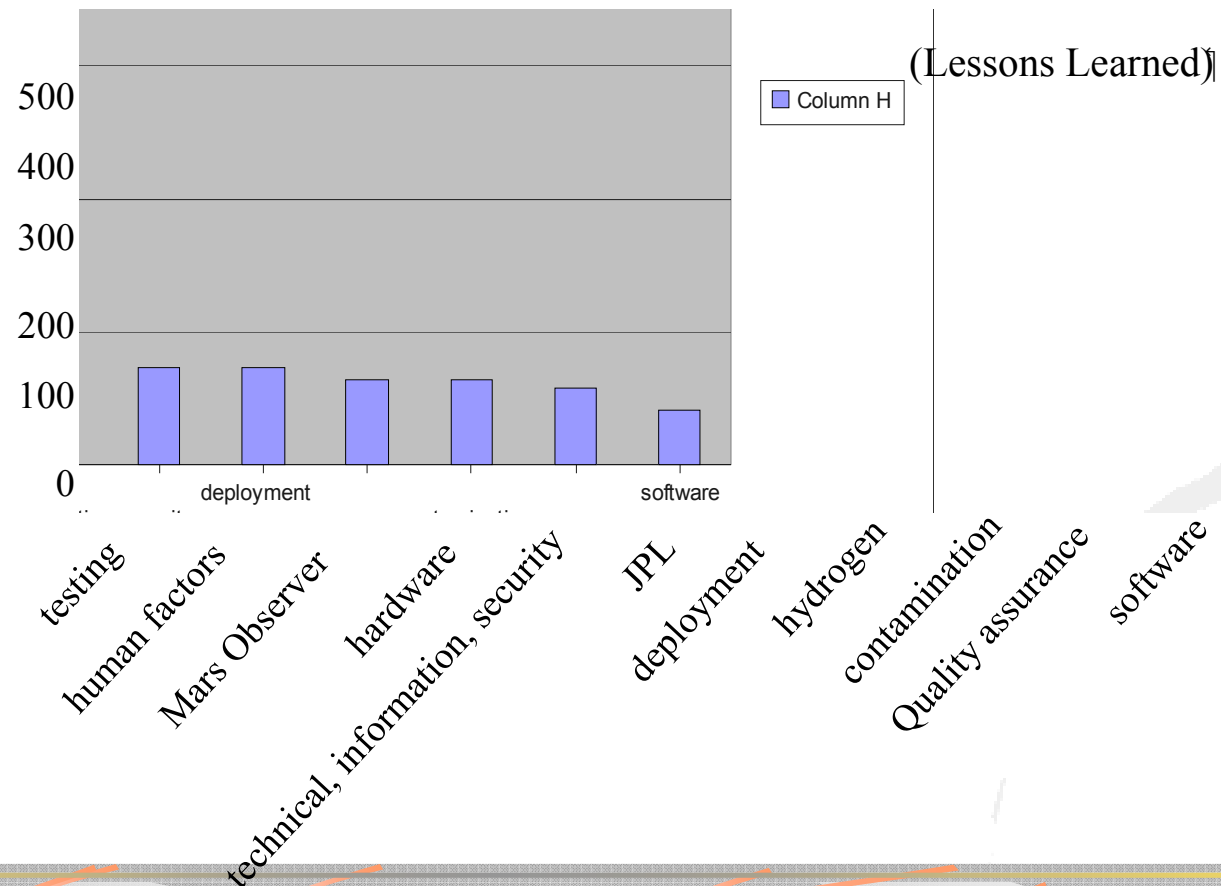- Policies
- Guidelines
- Test plans
- Designs

□ Column H

5k

4k

3k

2k

1k

0

human factors   Mars Observer

(Doc. #)   testing   Mars Observer

human factors

Queries in LLIS (after doc no.)

May be a good candidate for splitting

(Lessons Learned)

Column H

Axis	value
500	
400	
300	
200	
100	
0	

testing, human factors, Mars Observer, hardware, technical, information, security, JPL, deployment, hydrogen, contamination, Quality assurance, software
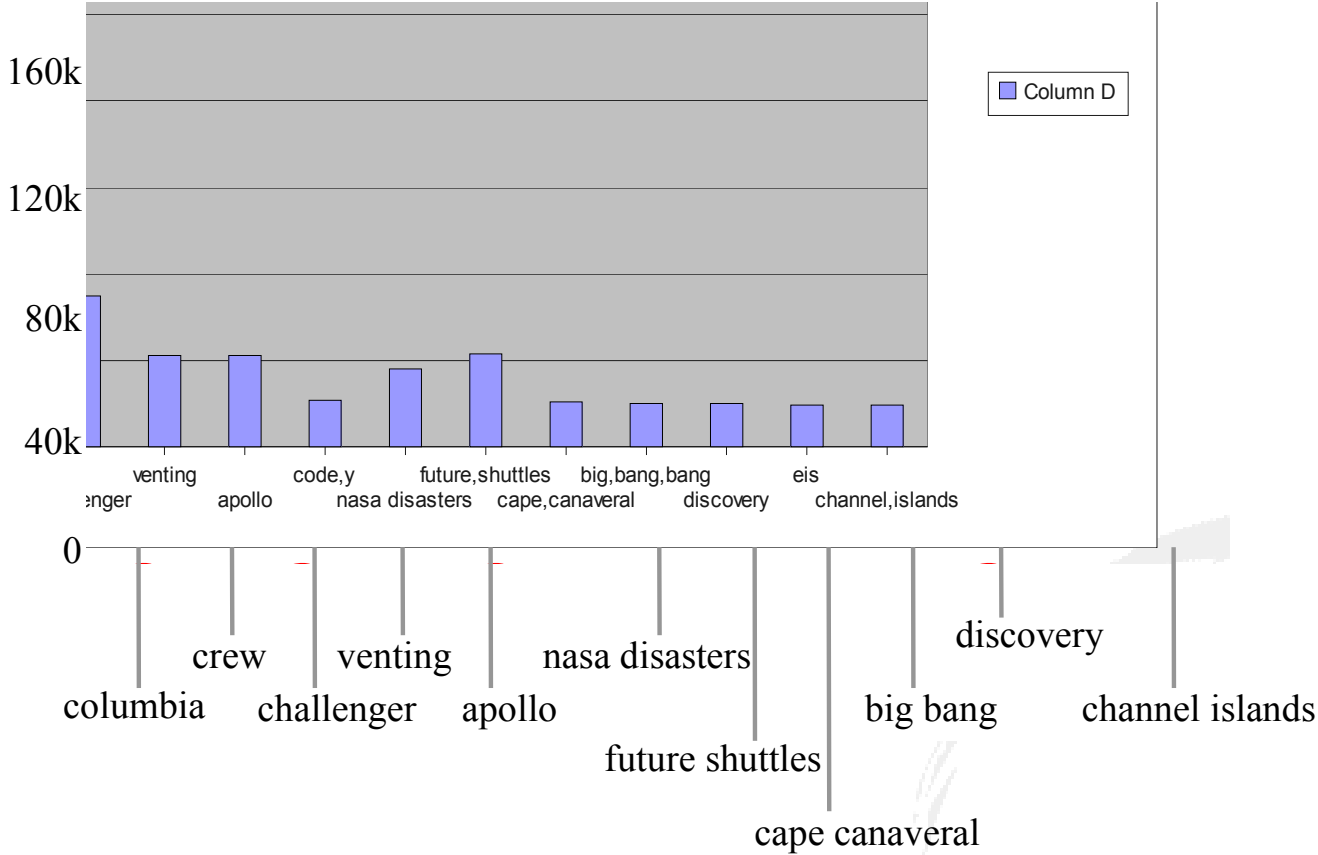
deployment   software

Top General Queries

Query Term Co-occurrences

- Crews
- Photos
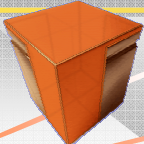- Costs
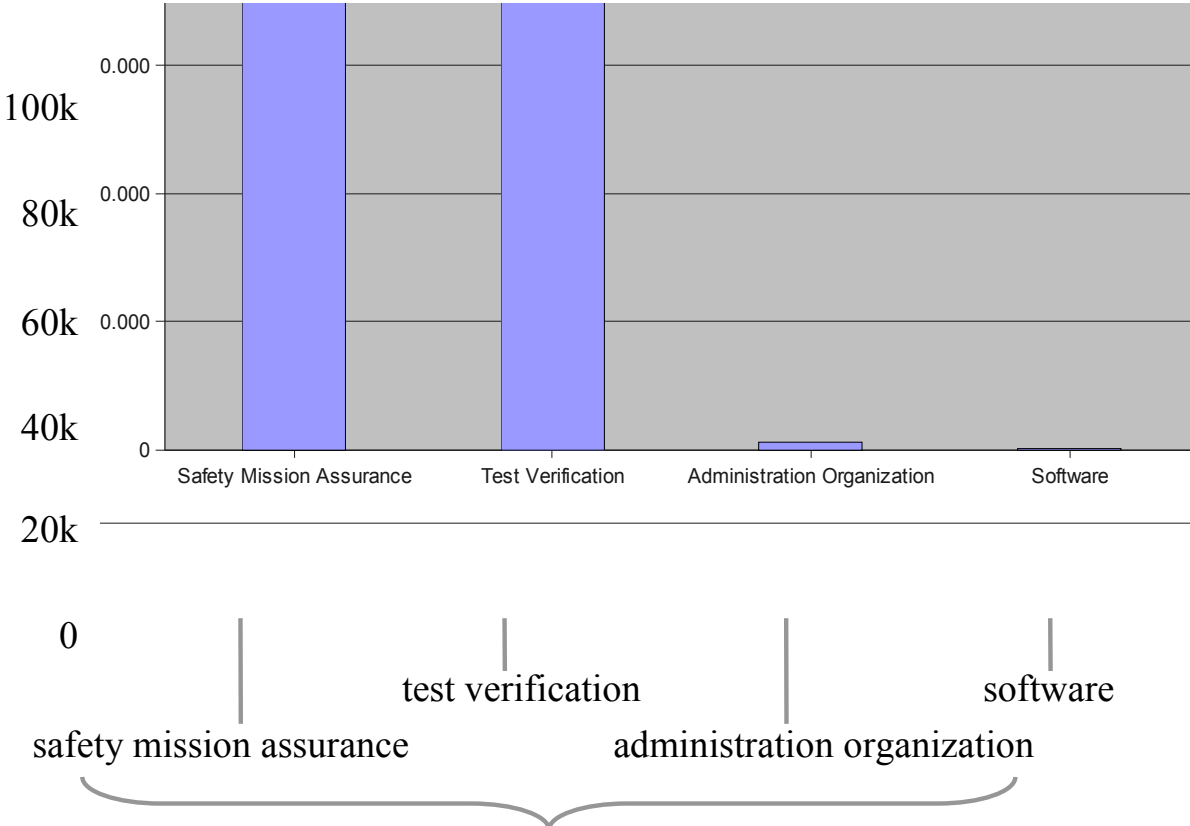- Payloads
- Blueprints
- Press Releases

Maintain compound terms or create new facets?

Clicks on Topics

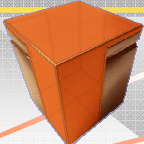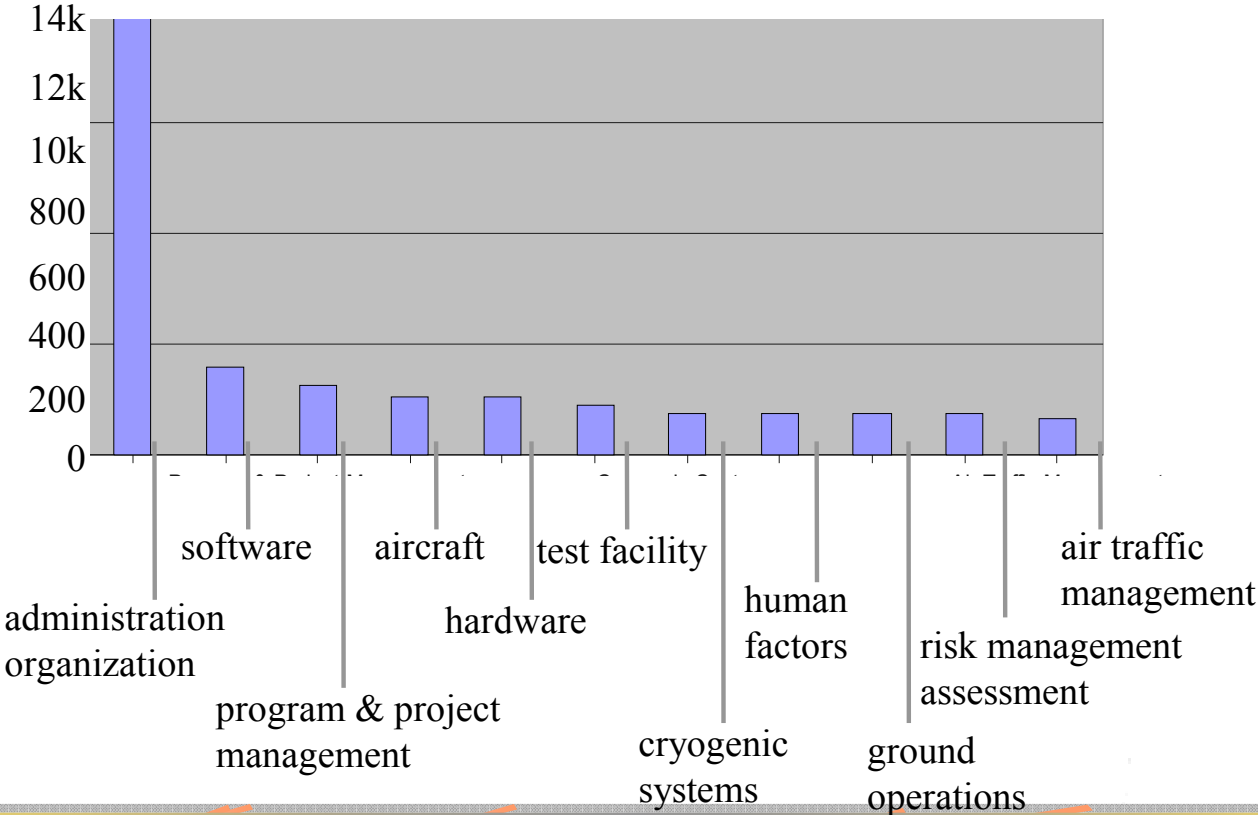Should be analyzed and mapped to the NASA taxonomy

May be good candidates for splitting

Clicks on Topics

after 'Safety Mission Assurance' and 'Test Verification'

## Should be analyzed and mapped to the NASA taxonomy

Recommendations

- Weighting
  - *"Too many results don't match."*
  - *"The resources are not being found, only other documents referencing them."*
  - *"There are too many different types of results."*
    - Searcher can't figure out how the engine parses their queries

- Indexing
  - "Where are the titles coming from?"
   - "NASA Software Quality Assurance Audits Guidebook" is indexed with a title of "SMAP-GB-A301"

Recommendations

- ## New Categories

  - Document Type - A popular access point to NEN content

  - Year - Individual years for more useful granularity (now grouped by decade)

- ## Stemming

  - Delimiters: [Underscores | Hyphens | Spaces]

 - Match to category values (i.e., Document Types)
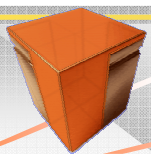
Recommendations

- Equivalencies (*Document Types* terms)
  - Document type synonyms
    - (such as: STD, NPR)
  - Document numbers (for popular documents)
   - (such as: 7120.2, 8739.8)
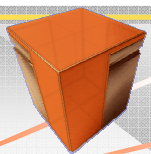
Recommendations

- Adjust Weights - Reduce the noise of full-text search
  - Favor controlled fields (particularly the Title field)
  - Minimize uncontrolled fields/document sections

- Search Syntax
 - Make it consistent & simple
  - Post the rules
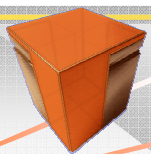
Recommendations

- Formatting (Context)
  - How
    - Include search string in lists of results
    - Put search string in boldface font in lists of results
    - Post repository descriptions at top of webpages when searching inside collections
  - Why
    - Help users learn the algorithm
    - Identify relevant items – QUICKLY!
   - Confidence in results
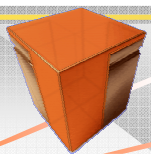
Recommendations

- Term Mapping
  - Topics (Verity) <---> Subject Categories (NASA Taxonomy)

- Splitting (Topics) (slide 22)
  - Safety Mission Assurance
  - Test Verification
  - Administration Organizations
  - Software

- Repositories - Candidates for indexing efforts
  - NACA sub-collection
 - Lessons Learned (LLIS)
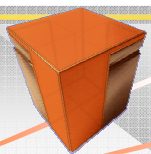
Recommendation

- Faceted Navigation
  - Good for cross-repository (25) search over many items
  - Lack of granularity in two categories
    - Could narrow unknown-item searches results
  - Browse & discovery
    - Framework for exploration and discovery within a universe of knowledge (a.k.a. research)
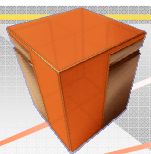
Methodology Recommendations

- New Metric: Relative Recall
  - Compares recall performance across search solutions over time
  - Re-use of data from Relevant Hits measure

- Relevancy Scale
  - Expand scale from 3-point to 5-point
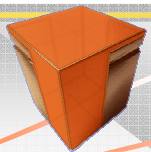
Methodology Recommendations

- Timeframe - Completed in 7 weeks

- Observations

  - Quantitative (10-?) and Qualitative (4-6)

  - Annually (bi-annually or quarterly until standard deviations are established)

  - Time-intensive

- Log Analyses

  - Monthly (recommended) or quarterly

  - Can be largely automated

- Surveys

  - Annually or bi-annually