Center of Excellence for Software Traceability (COEST) – An Update on TRACELAB+, plus Assessment of Provided Trace Matrices

Jane Hayes, PhD

hayes@cs.uky.edu

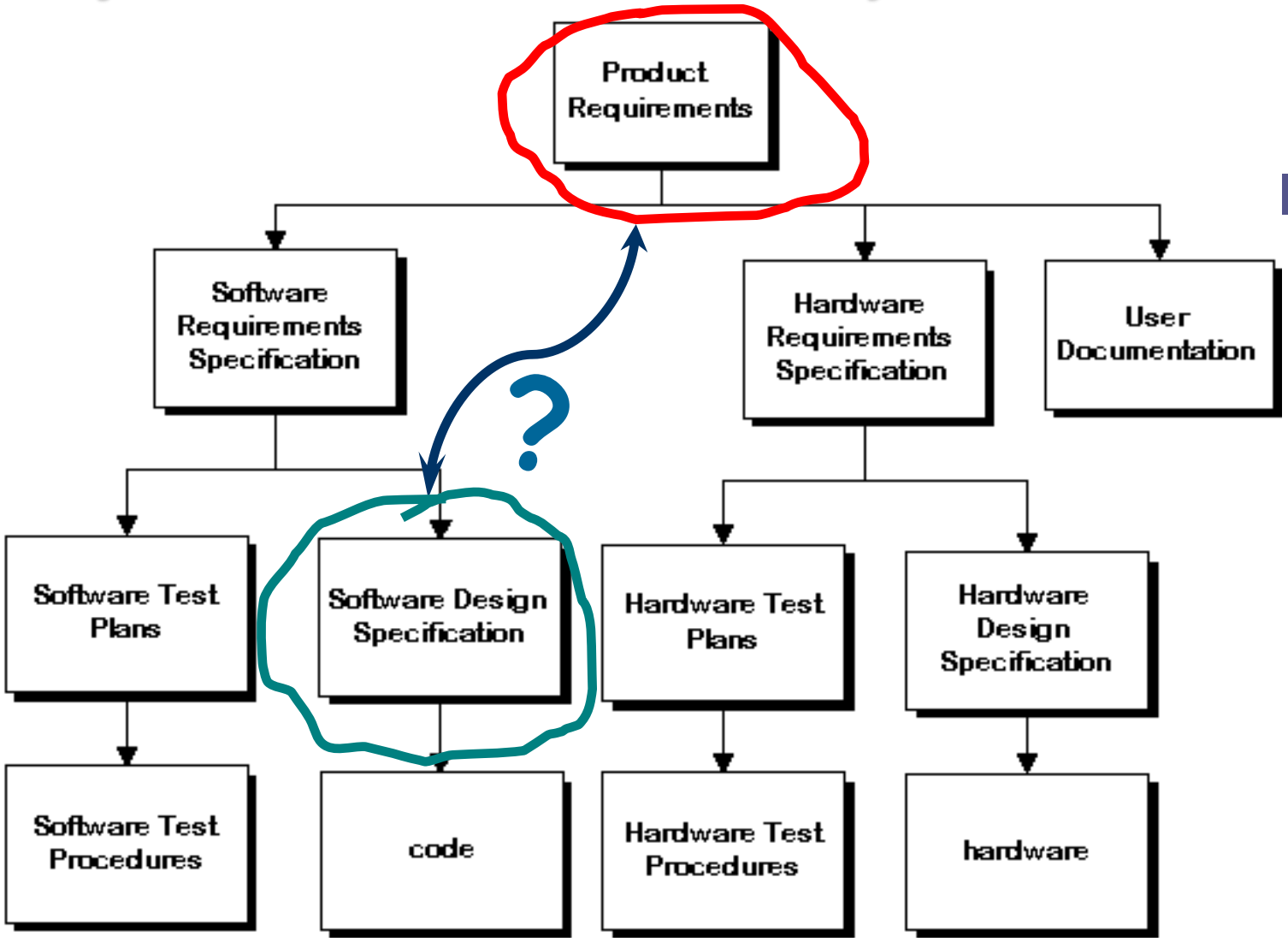
Funding for the work discussed in this presentation provided by the National Science Foundation grant CCF0810924 and NASA grant # NNX06AD02G .

Software Traceability

"*The ability to describe and follow the life of a requirement, in both a forward and backward direction, e.g. from its origins, through its development and specification, to its subsequent deployment and use, and through periods of ongoing refinement and iteration in any of these phases*"

Gotel & Finkelstein

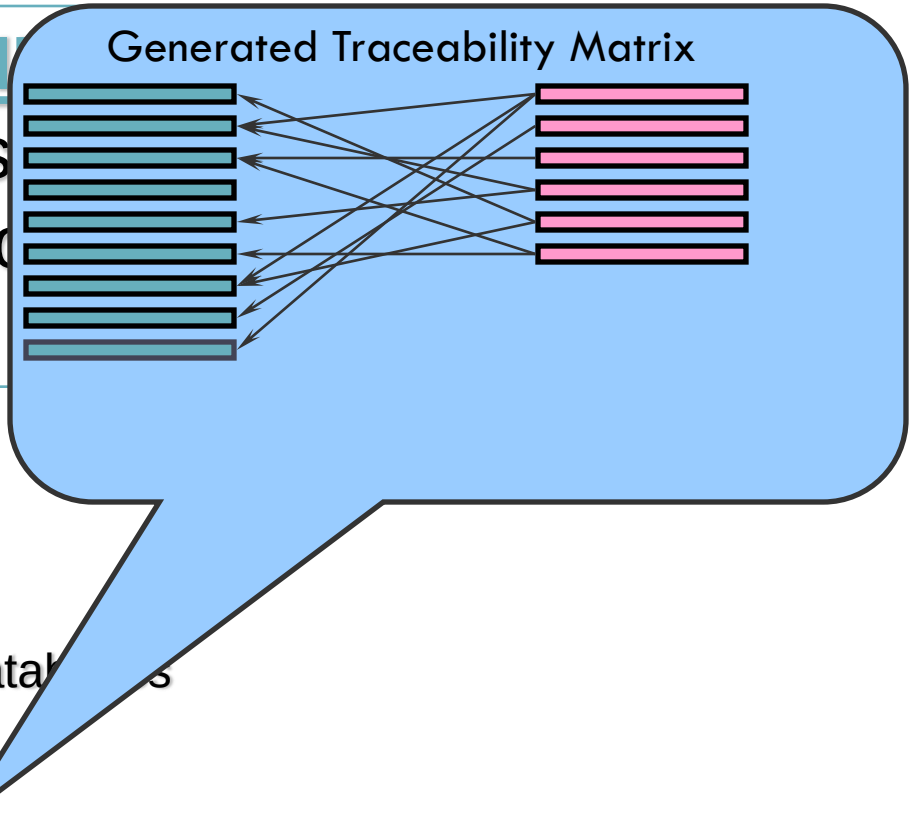
Project Document Hierarchy



- How do we verify that all requirements have been met?
- Is that all design elements exist to satisfy requirements?

Information Retrieval (to the rescue)

Information Retrieval (I
methods and techniques
relevant documents in d
given user queries.



- Intersection of Statistics, AI and Datal s
- Proven success

Importance of TMs

- Some Software Engineering Activities that Require TMs
  - Criticality analysis
  - Requirement satisfaction
  - Change impact analysis
  - Hazard reachability analysis
  - Regression testing
  - Traceability analysis
  - Risk analysis
  - Test Plan and Test Case Generation
  - Interface analysis
  - Consistency checking

How Ensure TMs are Accurate?

- Manual review
  - Time consuming
  - Error prone
  - Mundane
- Spot check
  - Incomplete
  - Mundane
- Assisted checking – enter Trace Matrix Analysis

Trace Matrix Analyzer (TMA)

- Treat TM as graph (edges, nodes)
- Apply graph heuristics to TM analysis
  - Children with too many parents?
  - Parent with no child?
  - Possible missing links?
  - Possible bad link?
- Visualize possible issues

Ideas based on Port, D., Hayes, J. Huffman, Huang, LiGuo, Nikora, Allen, "Text Mining Support for Software Requirements: Traceability Assurance," in Proceedings of IEEE Computer Society Hawaii International Conference on System Sciences (HICSS), January 2011, HICSS 2011: 1 – 11.

Advantages

- Greatly reduce workload of those performing requirements assurance
- Simple to operate
- Friendly User Interface (UI)
- Powerful expandability

Ideas based on Port, D., Hayes, J. Huffman, Huang, LiGuo, Nikora, Allen, "Text Mining Support for Software Requirements: Traceability Assurance," in Proceedings of IEEE Computer Society Hawaii International Conference on System Sciences (HICSS), January 2011, HICSS 2011: 1 – 11.

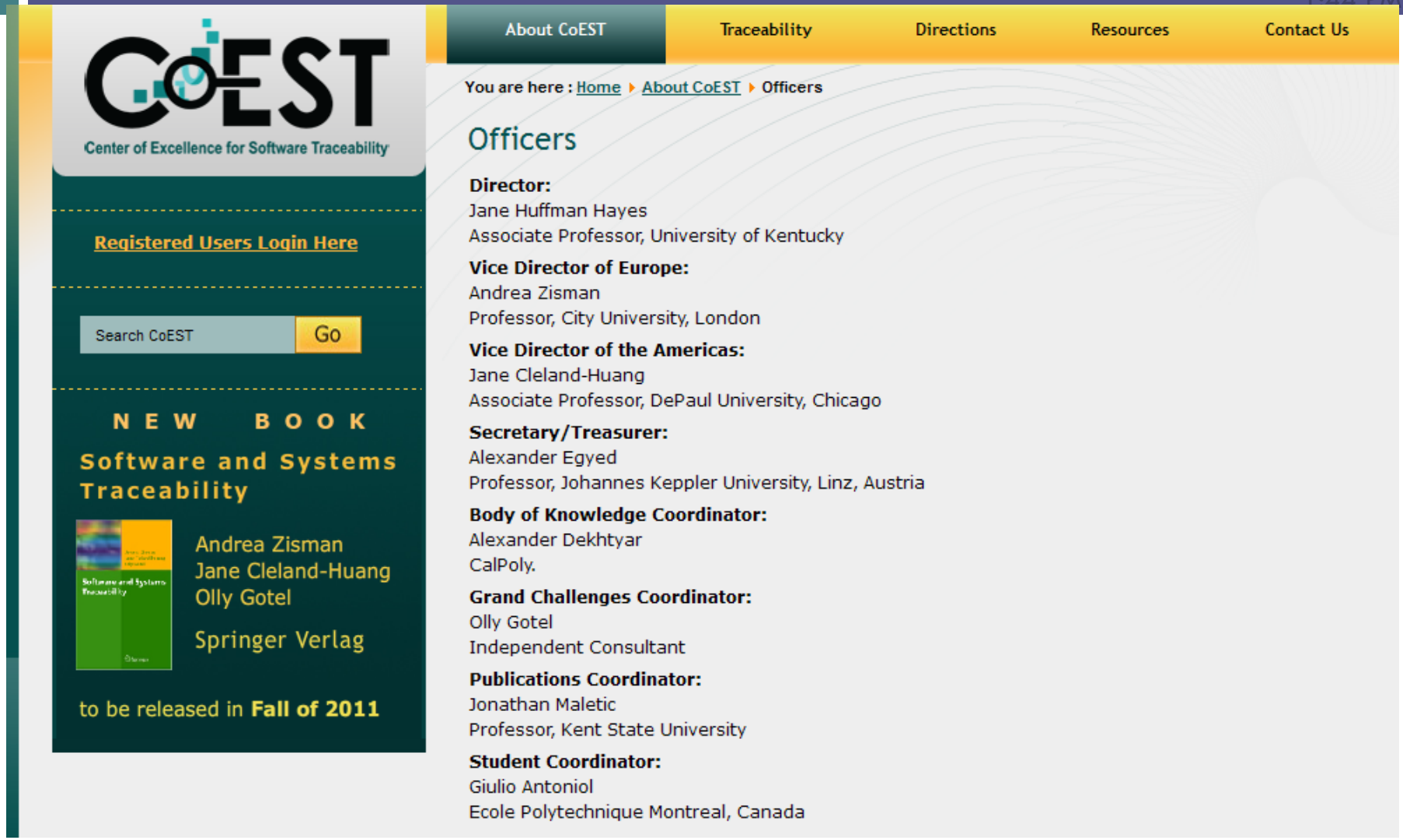
DEMO of TMA

CoEST's vision

- The **vision** of the COE for Software Traceability is to provide leadership for traceability research, education, and practice; promoting the pursuit of excellence from research idea to practice, based on a foundation of innovative, ethical, collaborative work
- Seed funding was provided by NASA and NSF

Everyone is welcome to join!

COEST Organization



Ubiquitous Traceability

- **Major Research Project:** RP1.1 Provide automation such that traceability is encompassed within broader software and systems engineering processes, and is integral to all tool support
- **Supporting Research Projects:** RP1.2 Embed traceability into all the software and systems engineering techniques and methods that it facilitates, and transfer this into industrial tool support
- RP1.3 Total automation of trace creation and trace maintenance, with quality and performance levels superior to manual efforts

Benchmarks

- A benchmark is a point of reference by which something can be measured
  - A program that is specially designed to provide measurements for a particular operating system or application
  - A set of performance criteria which a product is expected to meet
  - A set of conditions against which a product or system is measured

Benchmarks

- **Define a task**
  - Retrieve/Generate traces from high level to low level requirements
- **Provide datasets**
  - CM1, HIPAA to World Vista, IBS
- **Agree on a core set of metrics**
  - Recall, Precision, Lag, Average Precision
- **Capture/Report benchmarked results**

TraceLab- The Vision

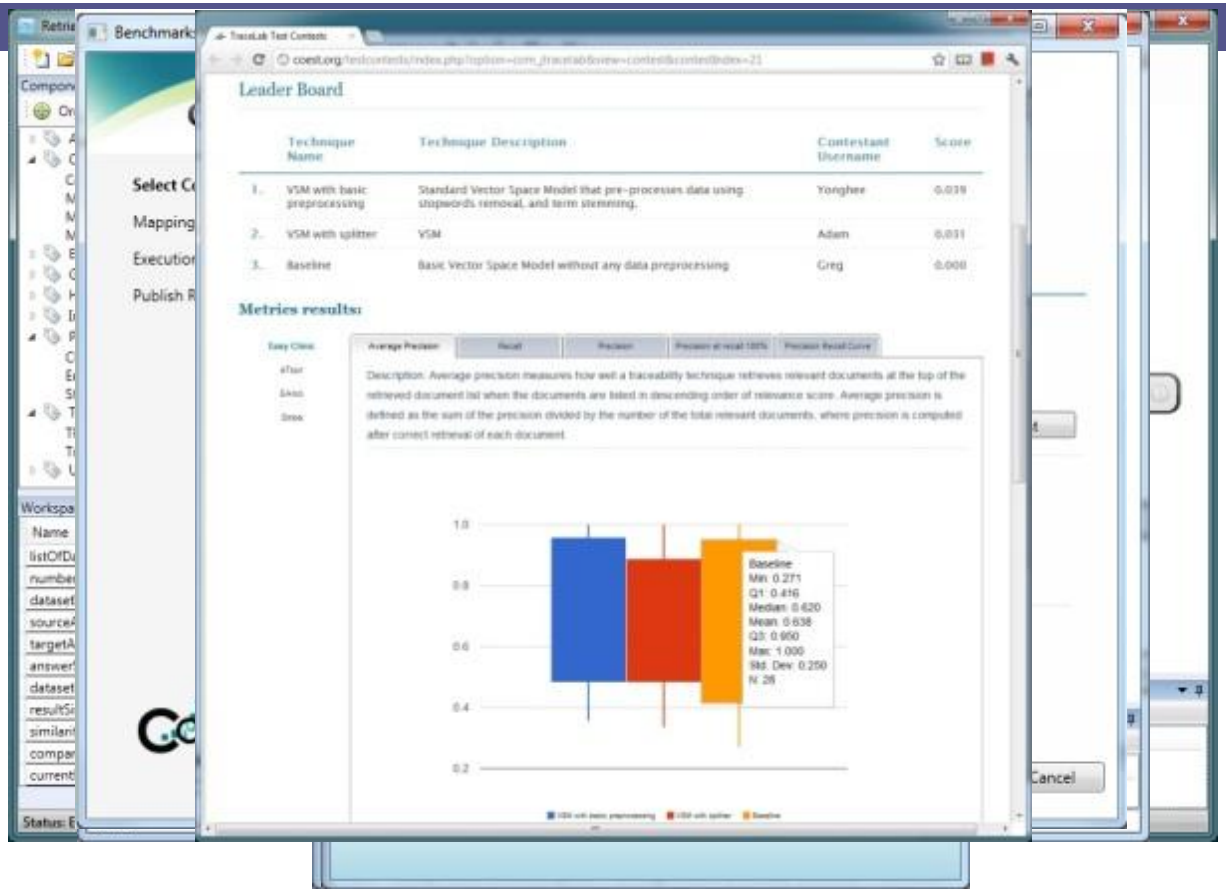
- Build a tool, similar to MatLab, but designed specifically for the traceability community
- Equip new researchers with basic algorithms and components
- Make it easier to perform rigorous comparative evaluations
  - Datasets
  - Benchmarks
  - Repeatable experiments
- Permit practitioners to use “best” algorithms for specific benchmark

TraceLab

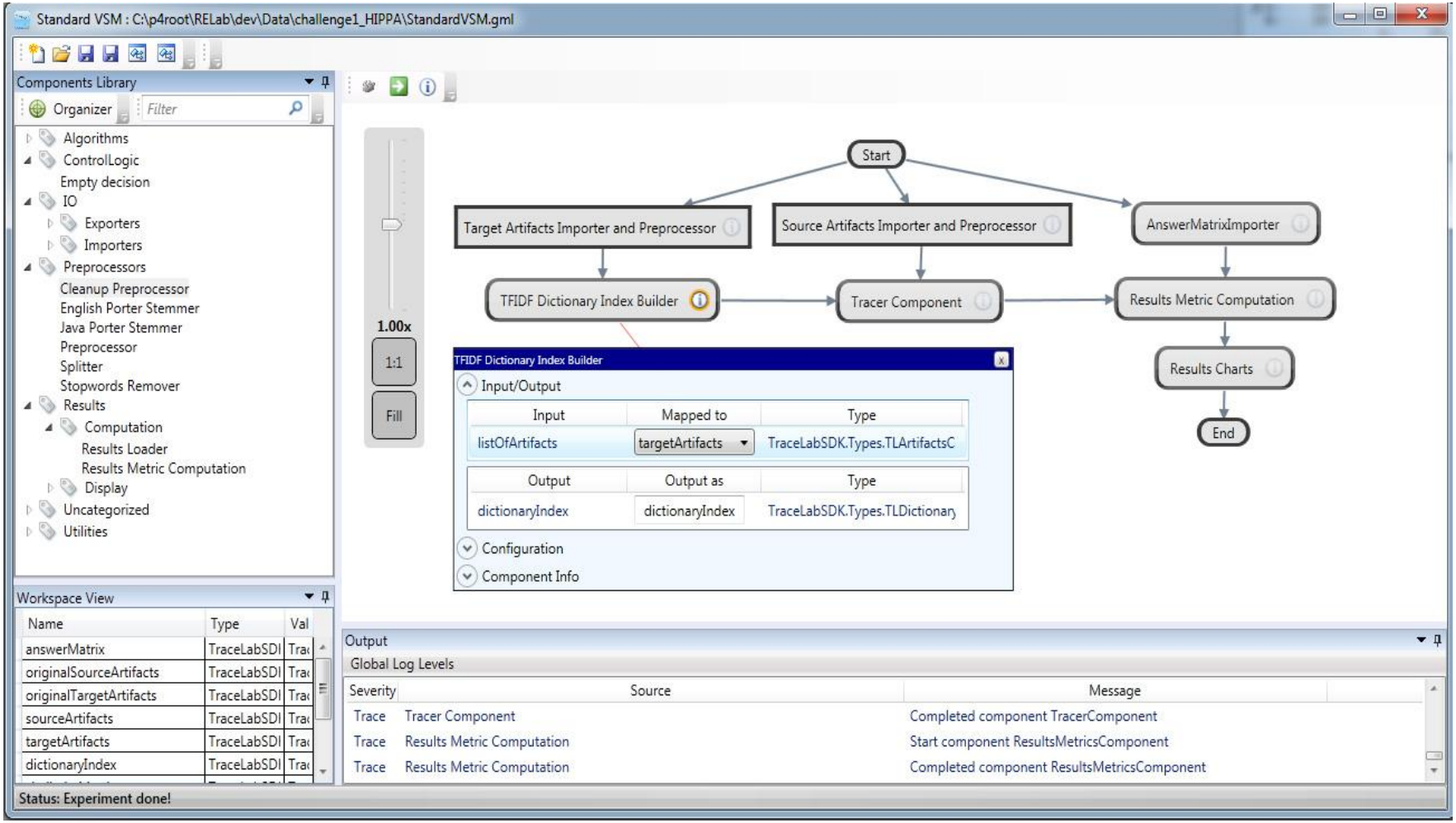
- Research environment designed to allow researchers to visually compose, execute traceability experiments using library of shared components
- Components in any memory managed language such as Java, C#, etc. TraceLab also allows calls to tools such as Matlab, R, etc.
- TraceLab currently runs in Windows environment but designed to port to Linux

TraceLab - The Role of Contests

- Define contests for community participation
  - Task (feature location)
  - Data set (benchmark)
  - Collection of “frozen” components with one “open”
- Prize to winner
- Permit practitioners to use “best” algorithms for specific benchmark (“player” from contest winner)


Benchmark
Leader Board
Technique Name
Technique Description
Contestant Username
Score
1. VSM with basic preprocessing
Standard Vector Space Model that pre-processes data using stopwords removal, and term stemming.
Yonghee
0.039
2. VSM with splitter
VSM
Adam
0.031
3. Baseline
Basic Vector Space Model without any data preprocessing
Greg
0.000
Metrics results:
Average Precision
Recall
Precision
Precision at recall 100%
Precision Recall Curve
Description: Average precision measures how well a traceability technique retrieves relevant documents at the top of the retrieved document list when the documents are listed in descending order of relevance score. Average precision is defined as the sum of the precision divided by the number of the total relevant documents, where precision is computed after correct retrieval of each document.
Baseline
Min: 0.271
Q1: 0.416
Median: 0.620
Mean: 0.638
Q3: 0.950
Max: 1.000
Std. Dev: 0.250
N: 28
Cancel

TraceLab Environment



Standardized Metrics

Evaluation Results

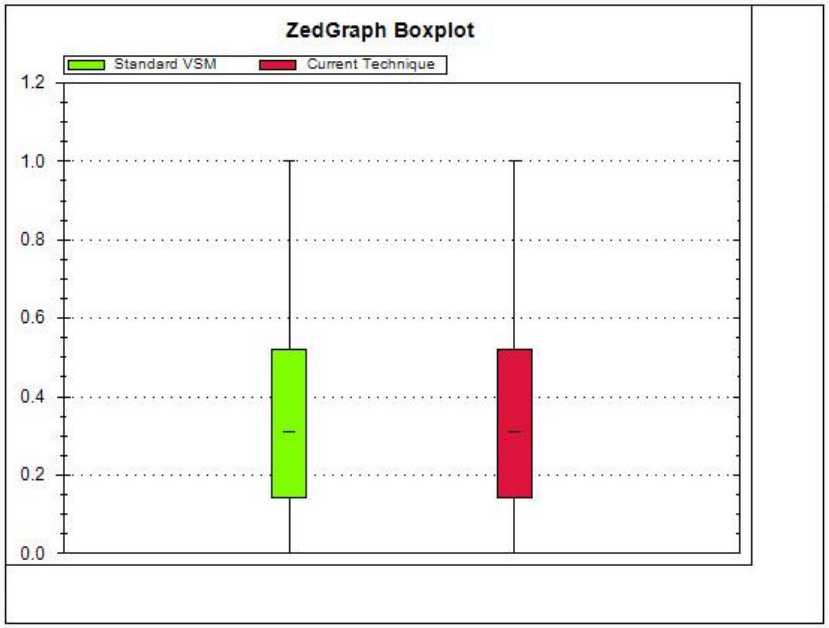
Evaluation Results

Below you can see the evaluation metrics of your results for each dataset.

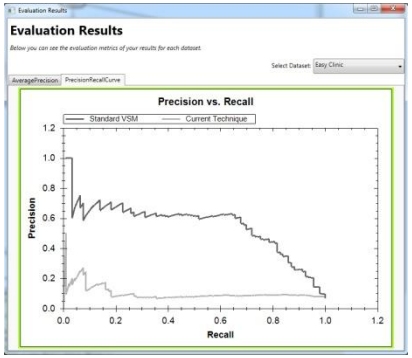
Select Dataset: CM1

CM1
Easy Clinic UC to TC

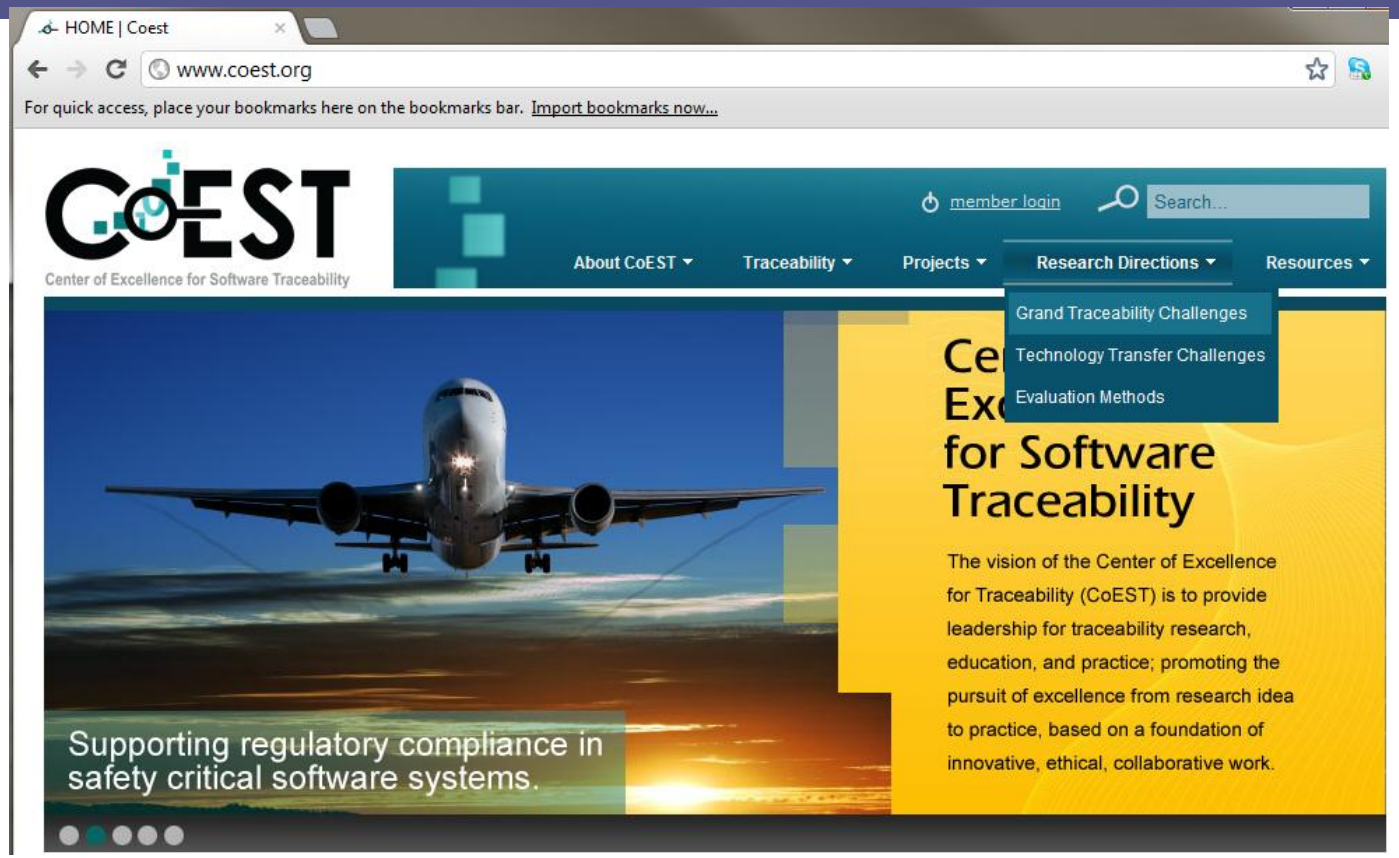
Recall | Precision | AveragePrecision



	Standard VSM	Current Technique
Min	0	0
Median	0.04	0.04
Mean	0.3692548...	0.36925486774382077
Max	1	1
Standard Deviation	0.29	0.29



Creating Contest for Traceability Techniques



Example Contests

- Contest 1
  - Task: Trace retrieval from use cases to code
  - Data Sets: EasyClinic, eTour, Eanci, SMOS
  - Metrics: Average Precision
- Contest 2
  - Task: Reducing human effort for relevance feedback
  - Data Sets: EasyClinic, eTour, Eanci, SMOS
  - Metrics: Average Precision, Number of feedbacks provided by human analyst

TraceLab Timeline

- Currently in beta-use at 6 universities
- Planned public release in July 2012 in conjunction with launching 5-6 research contests – culminating in The Grand Challenges of Traceability at ICSE 2013
- Will be open-sourced towards the Fall of 2012
- http://www.CoEST.org

DEMO of TRACELAB

Demo

Questions?

Backup

Challenges

- 2. **Purposed** - Traceability is fit for purpose and supports stakeholder needs.
- 3. **Cost-Effective** - The return from using traceability is adequate in relation to the outlay of establishing it.
- 4. **Configurable** - Traceability is established as specified, moment-to-moment, and the rich semantics accommodate changing stakeholder needs.
- 5. **Trusted** - All stakeholders have full confidence in the traceability, as it is created and maintained in the face of inconsistency, omissions and change; they can and do depend upon it.
- 6. **Scalable** - More and more artifacts are supported by traceability, of varying types and at variable levels of granularity, as the traceability extends through-life, and across organizational and business boundaries.
- 7. **Portable** - Traceability is exchanged, merged and reused across projects, organizations, domains, product lines and supporting tools.
- 8. **Valued** - Traceability is a strategic priority valued by all, where every stakeholder has a role to play and actively discharges his or her responsibilities.

What is a grand challenge?

What makes this a good Grand Challenge?

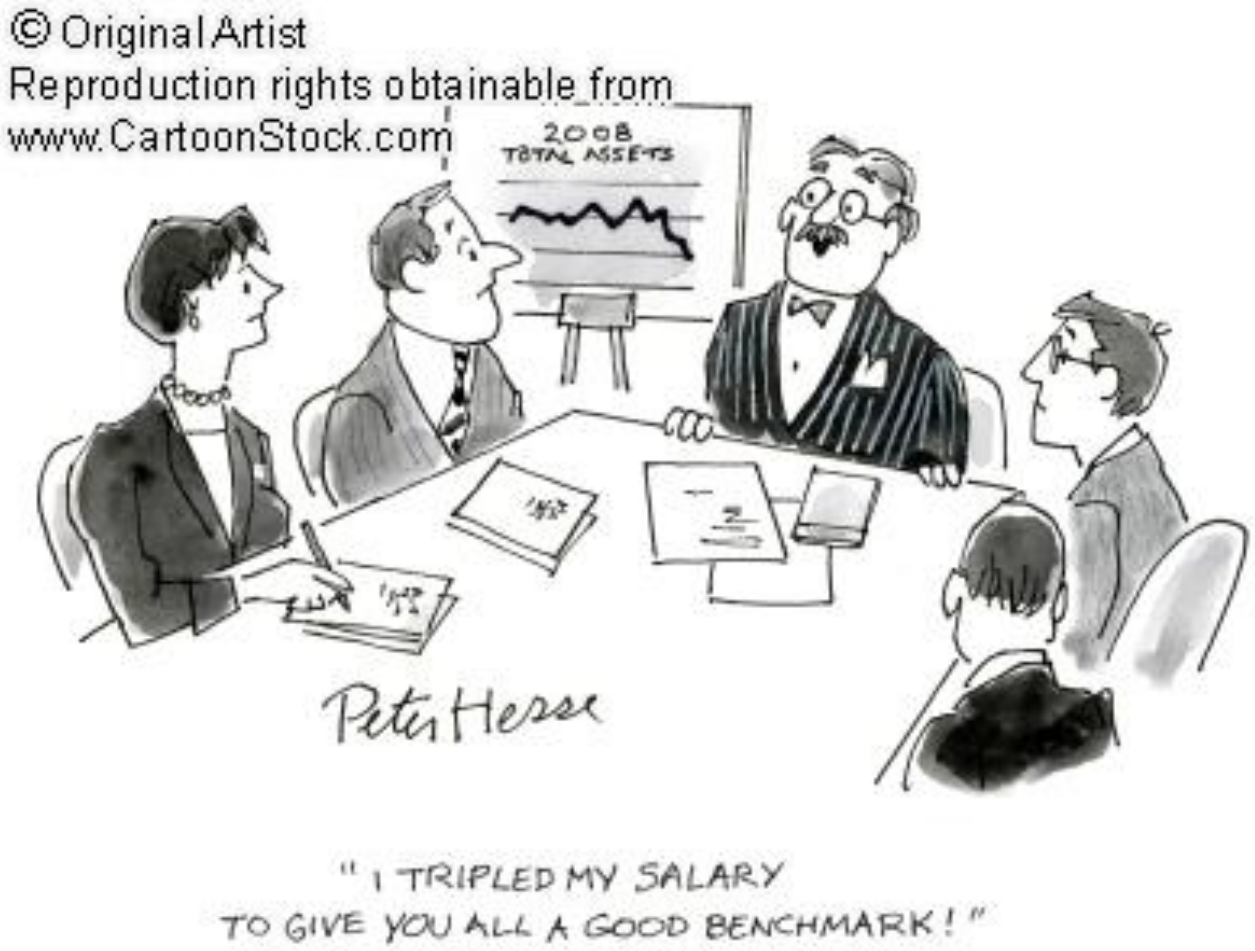
Is “Traceability” a grand challenge?

- Is traceability **important**? Why?
- Is traceability **difficult** to achieve?
- Do we have a **clear vision** of where we want to go?

Benchmarks

Recall vs. Precision problem – small changes in thresholds can have inordinate impact upon recall vs. precision – creating zigzag graphs. For benchmarking metrics, how do we overcome this?

Benchmarks



High water marks–

Will high benchmarks thwart innovation?

Is this a good or bad thing?

Benchmarks

Trust –
What kinds of checks and balances do we need to put into the process to make sure that benchmarks are fair?

How do we make comparisons anyway?

Benchmark issues

Early work

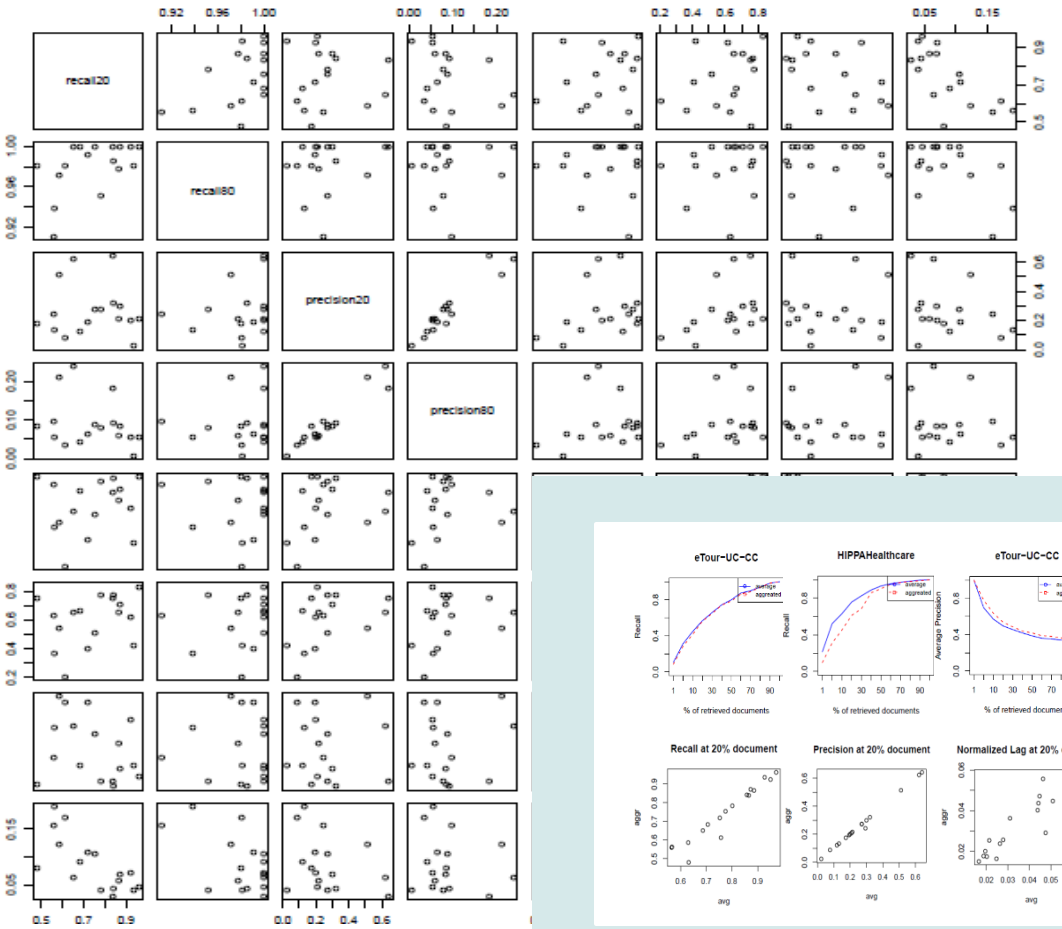
TEFSE community.

An idea.

Towards a grand challenge.

Beyond the challenges

TraceLab & Benchmarks



Yonghee's work

Benchmark insights

- What is the purpose of benchmarking our community?
  - What do we hope to accomplish from benchmarking?
- What are the major pitfalls of benchmarking in the traceability community?
  - How can we avoid them?