

Foundations of Next Generation Information Access

Warren Greiff

781-271-7421

greiff@mitre.org

Jay Ponte

781-271-8271

ponte@mitre.org

MITRE Sponsored Research



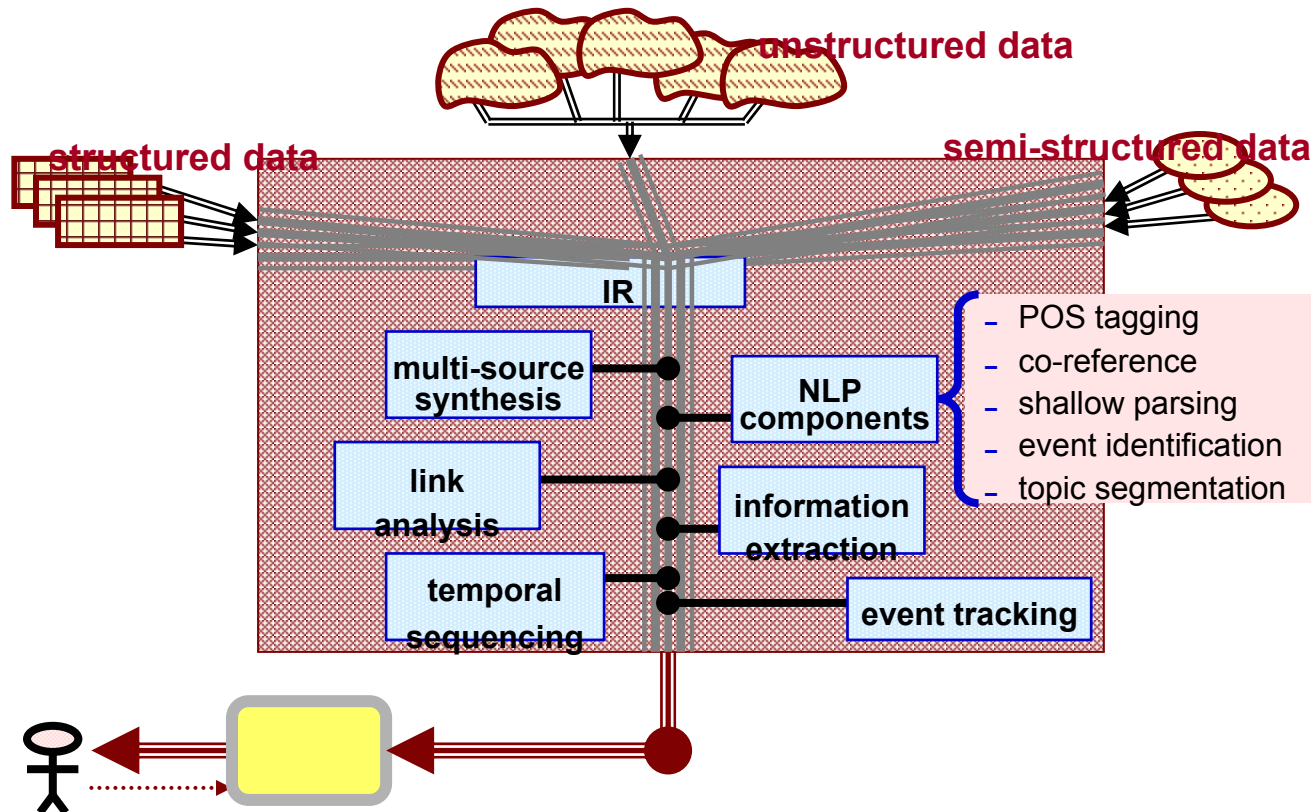
**MITRE
Technology
Program**

Problem

- **Overwhelming need for computerized support for information gathering**
 - particularly in areas of:
 - military readiness
 - terrorism response
 - disease control
- **Current technology for addressing problem:**
 - fragmented across research communities
 - integration is difficult

- ***Needed:***
 - *underlying formalism*
 - *cutting across different technologies*

Background



Next Generation Information Access
will require principled approaches to the
integration of multiple component technologies

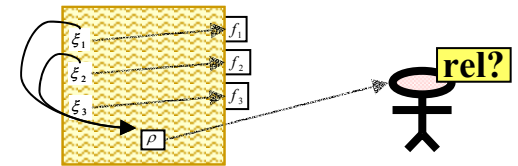
Objective

- **Statistical foundations for information access**
 - rigorous characterizations of:
 - probabilistic modeling
 - parameter estimation
- **Principled methodologies for adapting to**
 - new languages
 - new genres
 - new information domains
 - new tasks
 - auxiliary knowledge sources

Activities

■ Simulated stochastic modeling of:

- latent document features
- observable document features
- document relevance
- queries



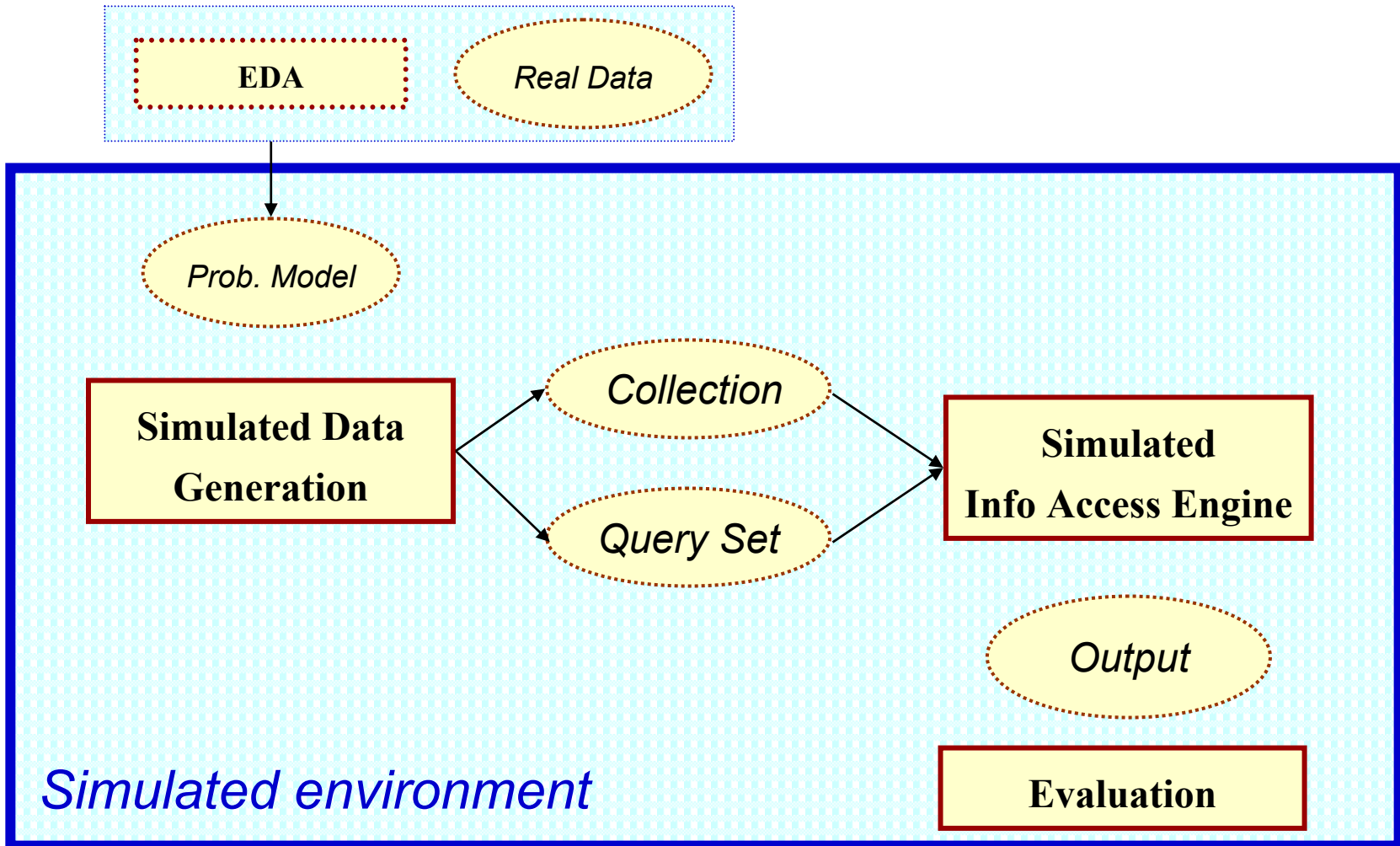
■ Exploratory data analysis

- available research corpora
- verification of simulation models

■ Central focus -- research into

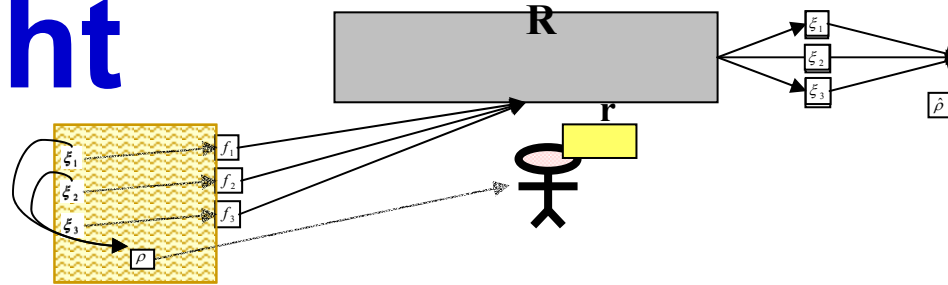
- variance reduction
- empirical Bayes techniques

Highlight

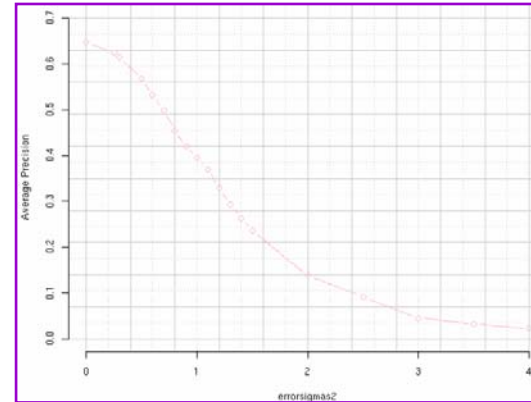
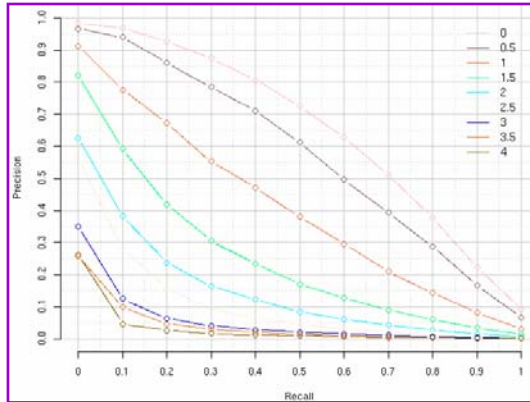


Environment for Information Access Simulation Studies

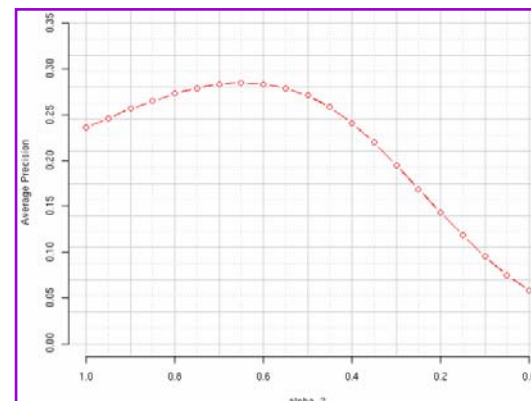
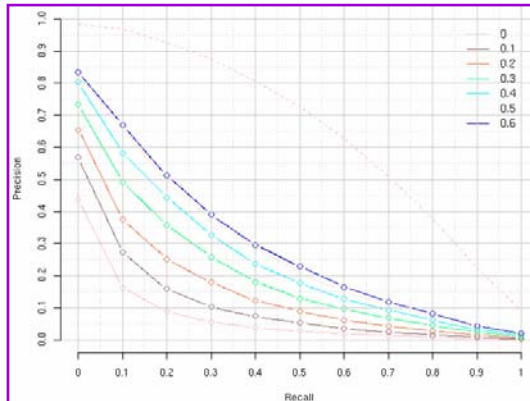
Highlight



adding noise



shrinkage



recall-precision curves

average precision

Improved retrieval performance achieved through application of estimator shrinkage

Impacts

- **Direct relevance to existing MITRE projects**
 - **GeoNode**
 - **TIDES**
 - **IDIA**
- **Information access systems development**
 - **incorporation of new sources of evidence**
 - **tailored to specific sponsor needs**
- **Scientific contribution**
- **Set Next Generation Info Access research agenda**
 - **Theoretical foundations**
 - **Catalyst architecture (DARPA TIDES)**

Future Plans

