Text Analysis Contributions to Augmented Cognition Work at Sandia

Travis Bauer

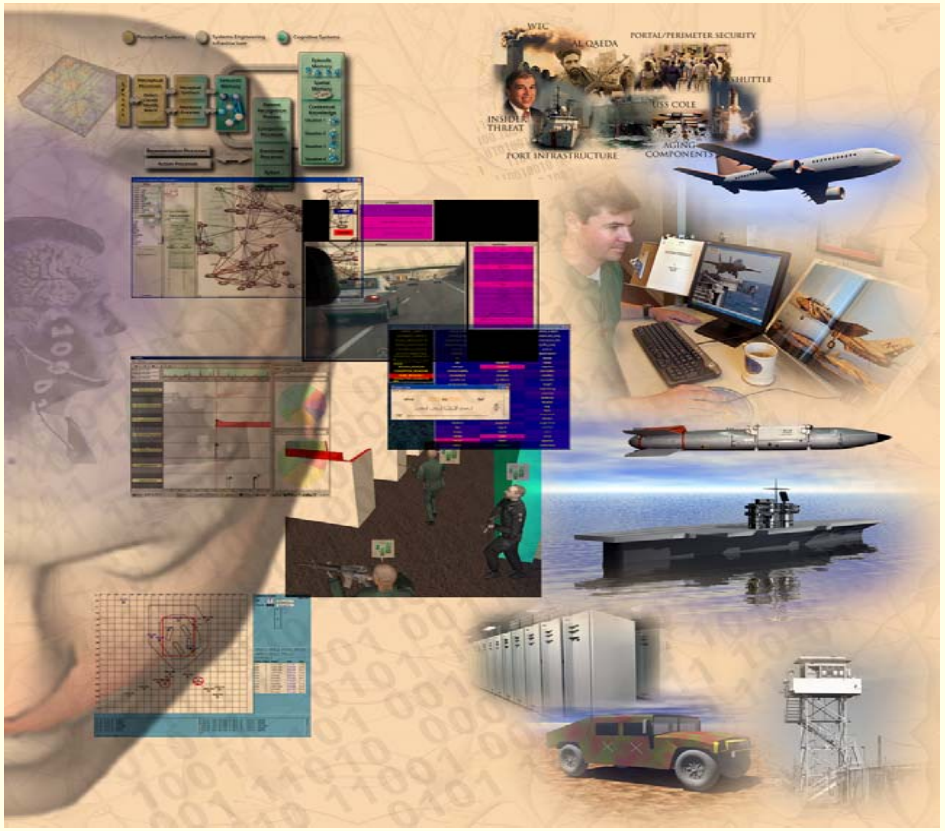
Cognitive and Exploratory Systems and Simulation

tlbauer@sandia.gov

284-8723

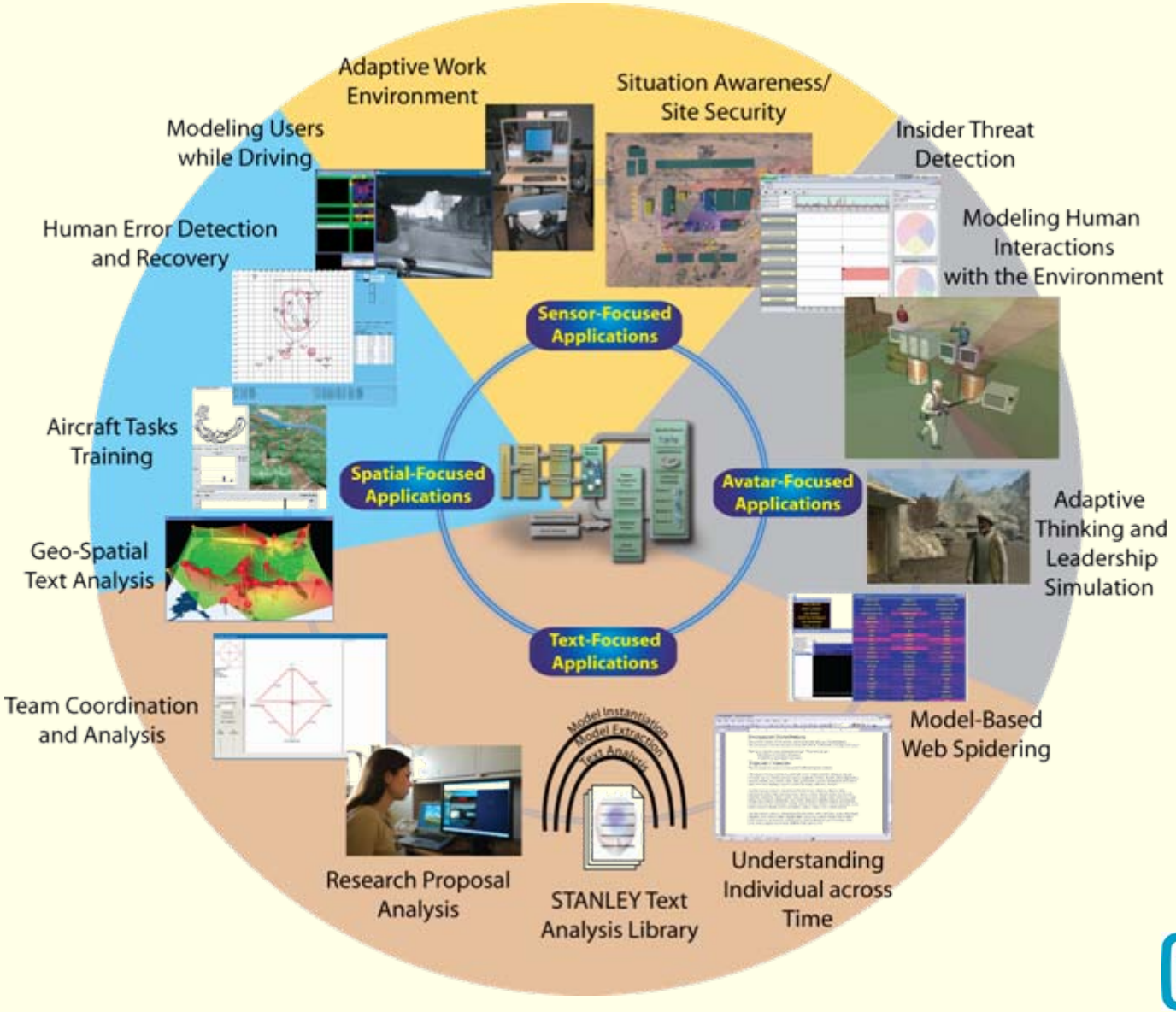
Sandia's Cognitive Systems and Technologies Program

Capture the <u>context</u> understanding of individual humans in technologies that augment human detection and interpretation of meaningful patterns.

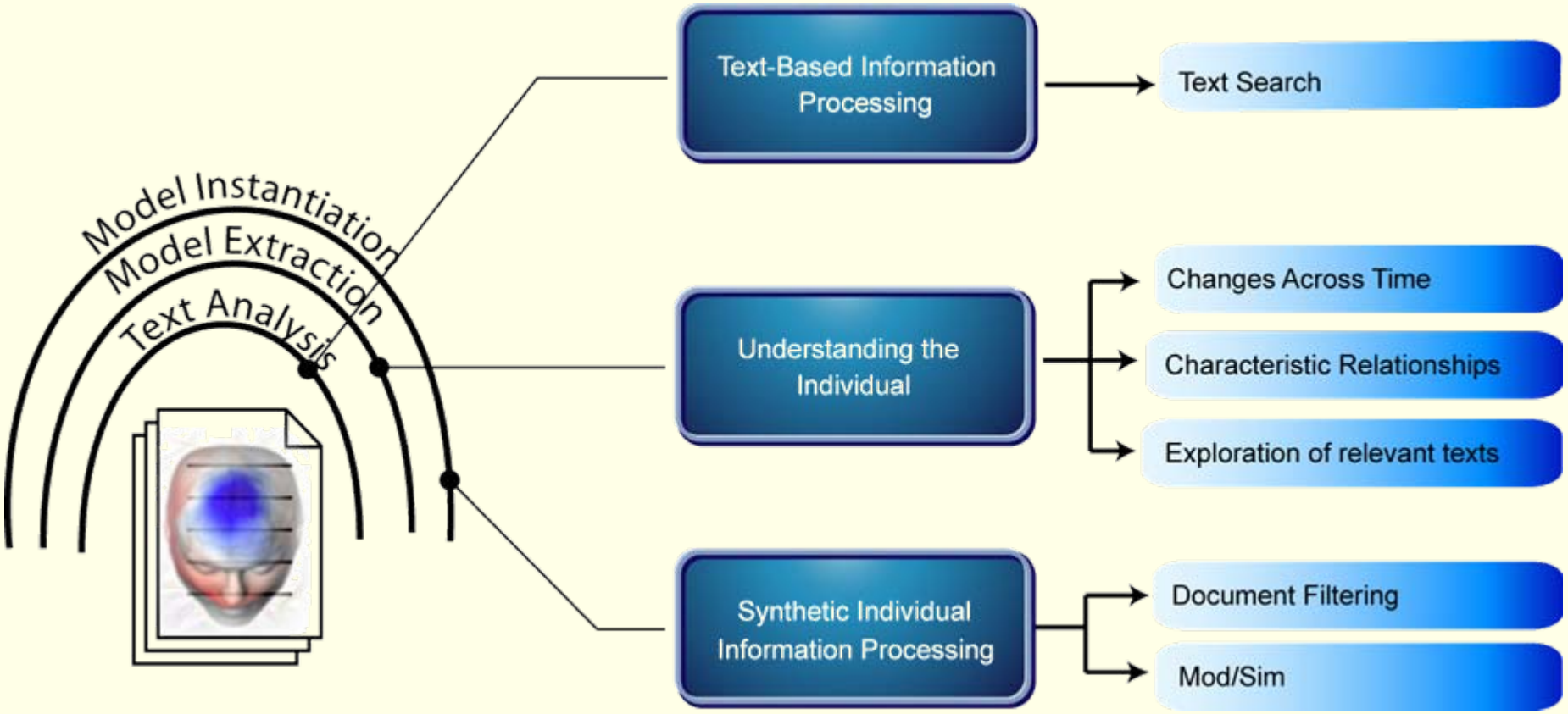


- **Program start 1999.**
- **Originally Internally Funding**
- **Sandia decided in 2005 to make Cognitive Systems an enduring Core Focus Technology of Laboratory**
- **External projects**
  - **DARPA, DOE, DaimlerChrysler, Honeywell and Boeing.**
  - **$11M (internal + external) total revenue in 05**
- **Early emphasis on core capability development, current emphasis on transition of capabilities to create technology solutions.**

AugCog Applications and Prototypes



Applying Text Analysis Techniques Lets us Automatically Build Models of Individuals for Various Applications



Problems We've Addressed with this Technology

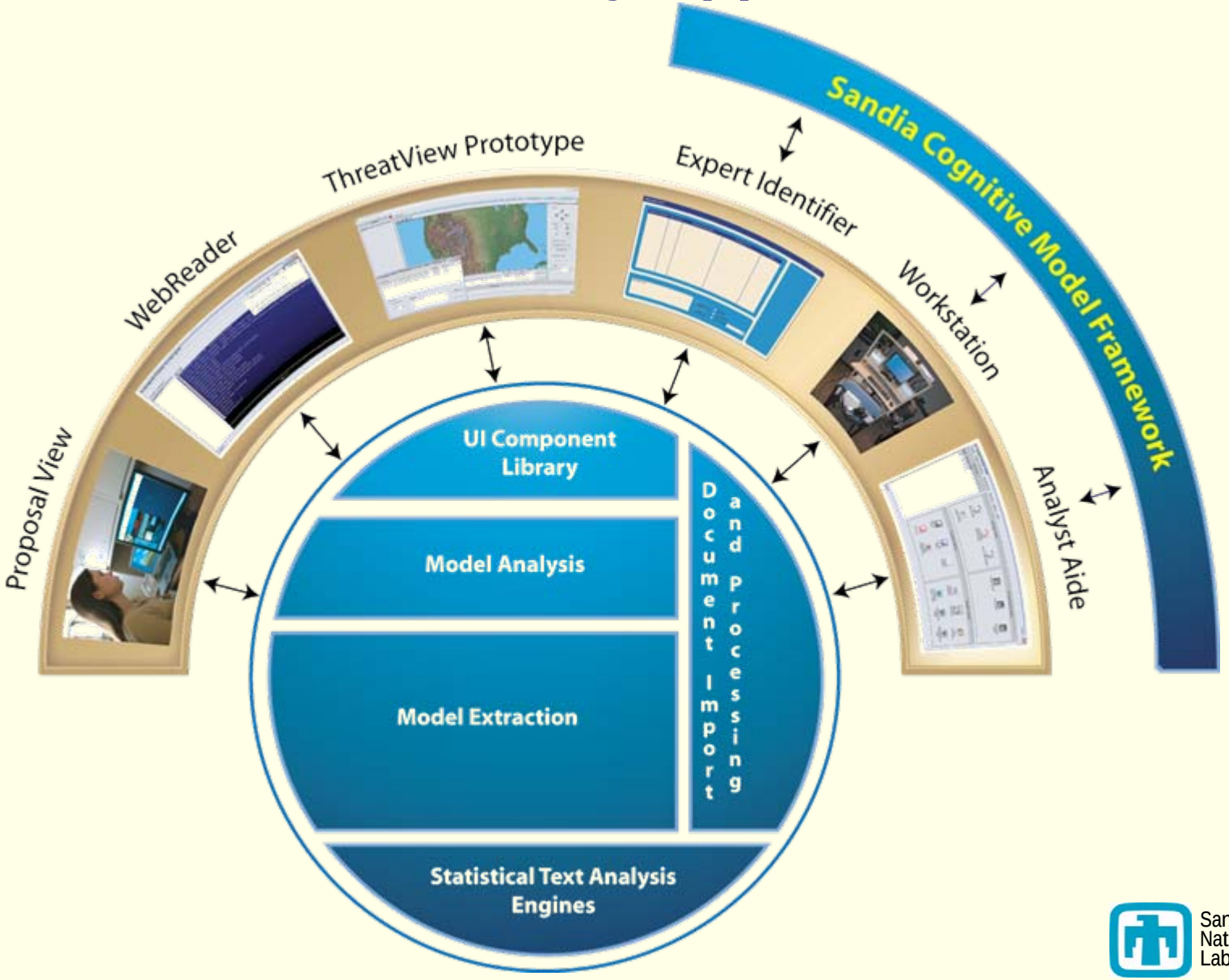
Statistical Text Processing

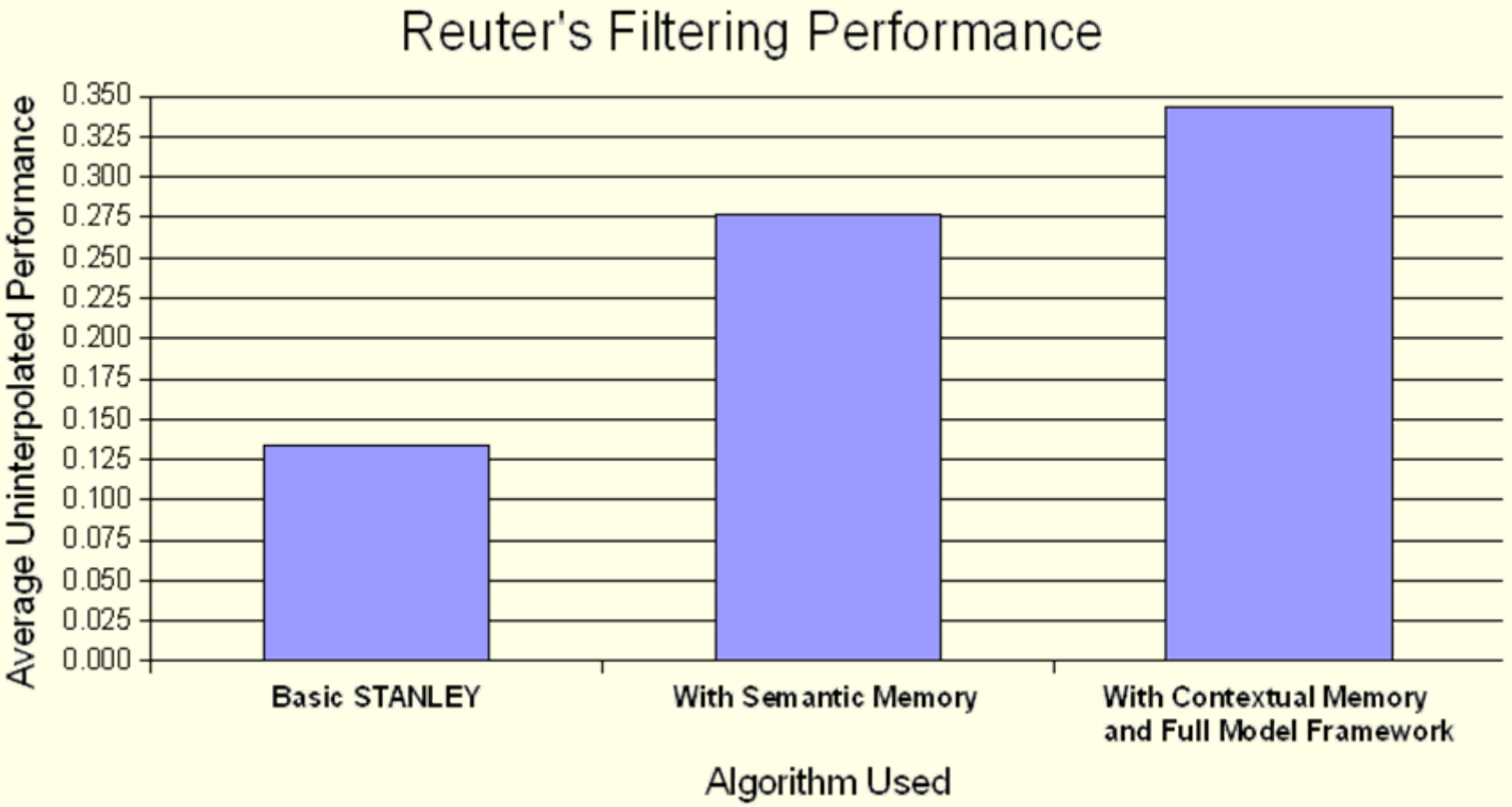
Model Extraction

Model Instantiation

Problem	Solution
Matching Internal Research Proposals to Calls	**Integrated STANLEY with Sandia's internal grant application process. Was available to all Sandians through the internal web.**
Helping analysts understand what requirements are relevant in certain circumstances.	**Integrated search capability wrapped into a custom application to assist requirements evaluations.**
What topics are covered in a particular set of interviews?	**Built a model of the interviews, extracted and visualized key terms and relationships among those terms in a map. Observe the clusters.**
What dates might be significant for a particular group, causing them to change their approach?	**Collected news stories produced by the group across a several month period. Generated a reporting detecting shifts in characteristic properties of their model.**
Where might a spokesperson have spoken "out of step" with the groups official story?	**Built a model of the rhetoric of the group from news stories. Had the model read the stories and identify the ones that fit in the least.**
Need to find blogs about a particular topic.	**Built a model of a high quality blog. Had the spider go out and find other similar documents, which were the blogs of interest.**
Want to look for documents produced by a particular group.	**Built a model of documents from the group's leader. Had the web spider surf the net looking for related documents.**

STANLEY Library Applications

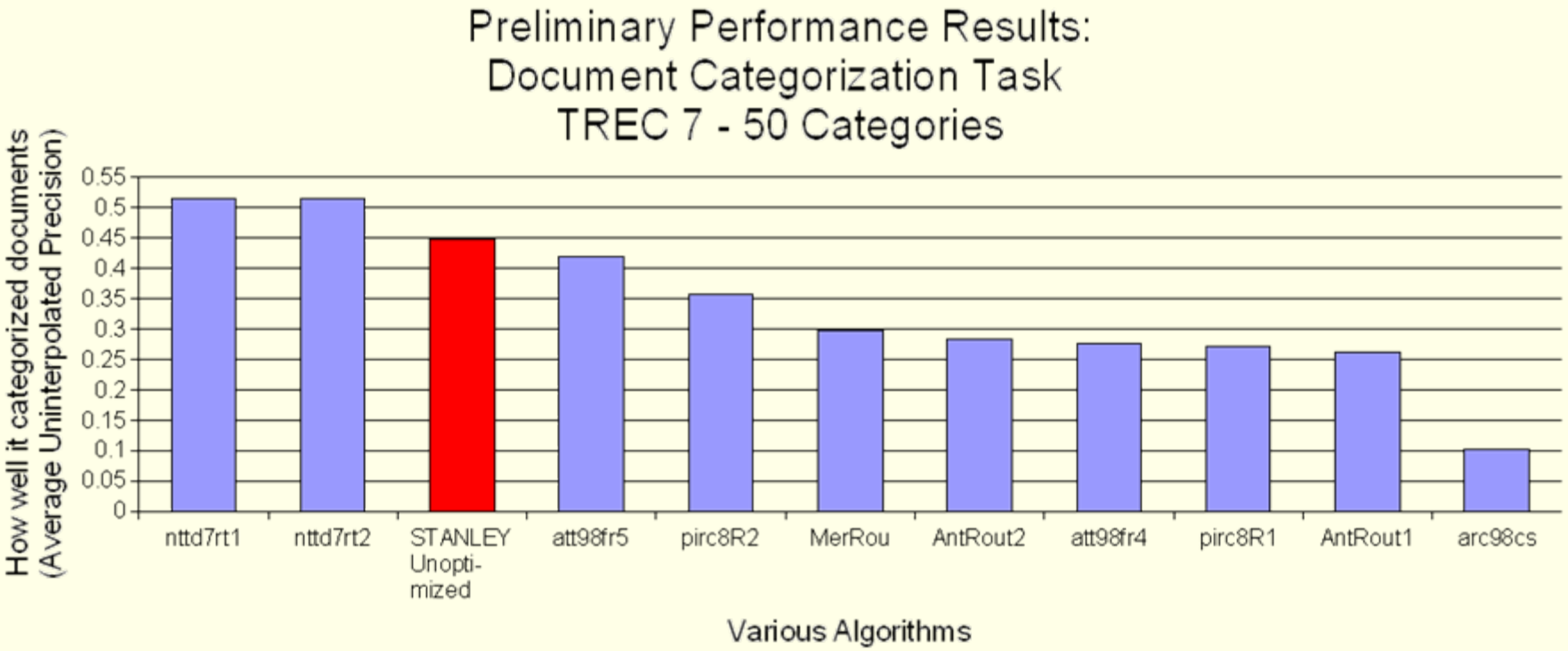




Filtering task created using the Reuters-21578 data set.

Filter the articles belonging to the topics "cocoa", "coffee", "gold", "lumber", "platinum", "rice", or "yen" from those articles that do not contain any topics.

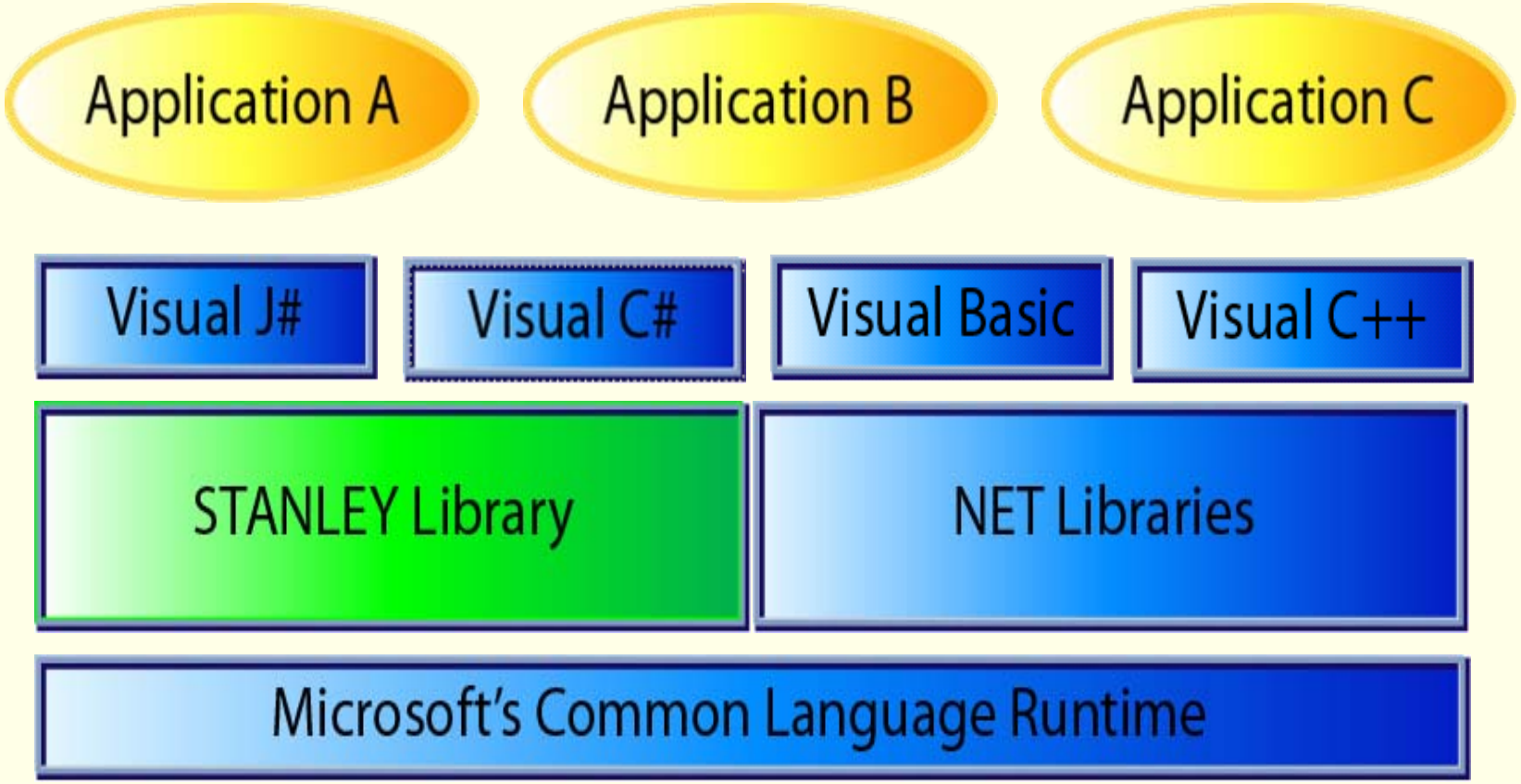
Model-based Document Discovery and Evaluation: Early Performance Results



In a standard information retrieval document categorization task, an unoptimized version of our model creation algorithm performed better than most published algorithms.

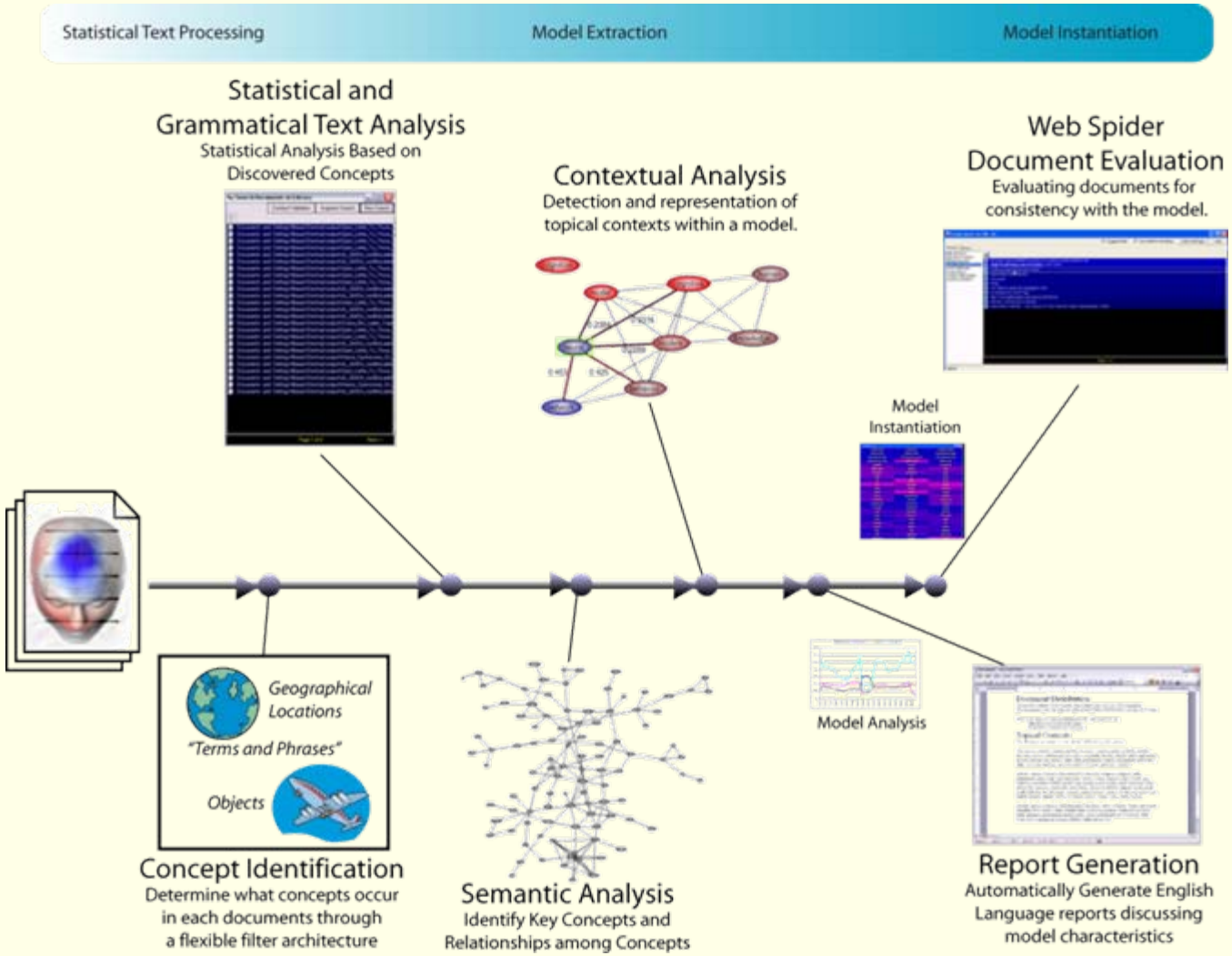
Backup Slides

STANLEY is built on the .NET platform



Cognitive Modeling through Text is . . .

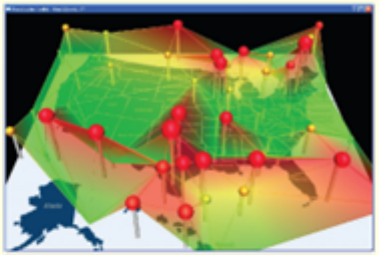
Automatically characterizing a person from text the produce or access



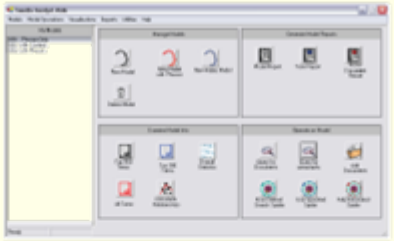
STANLEY is Engineered to Interoperate with Other Systems

Proposal View

ThreatView Prototype



Analyst Aide



Others

Web Server

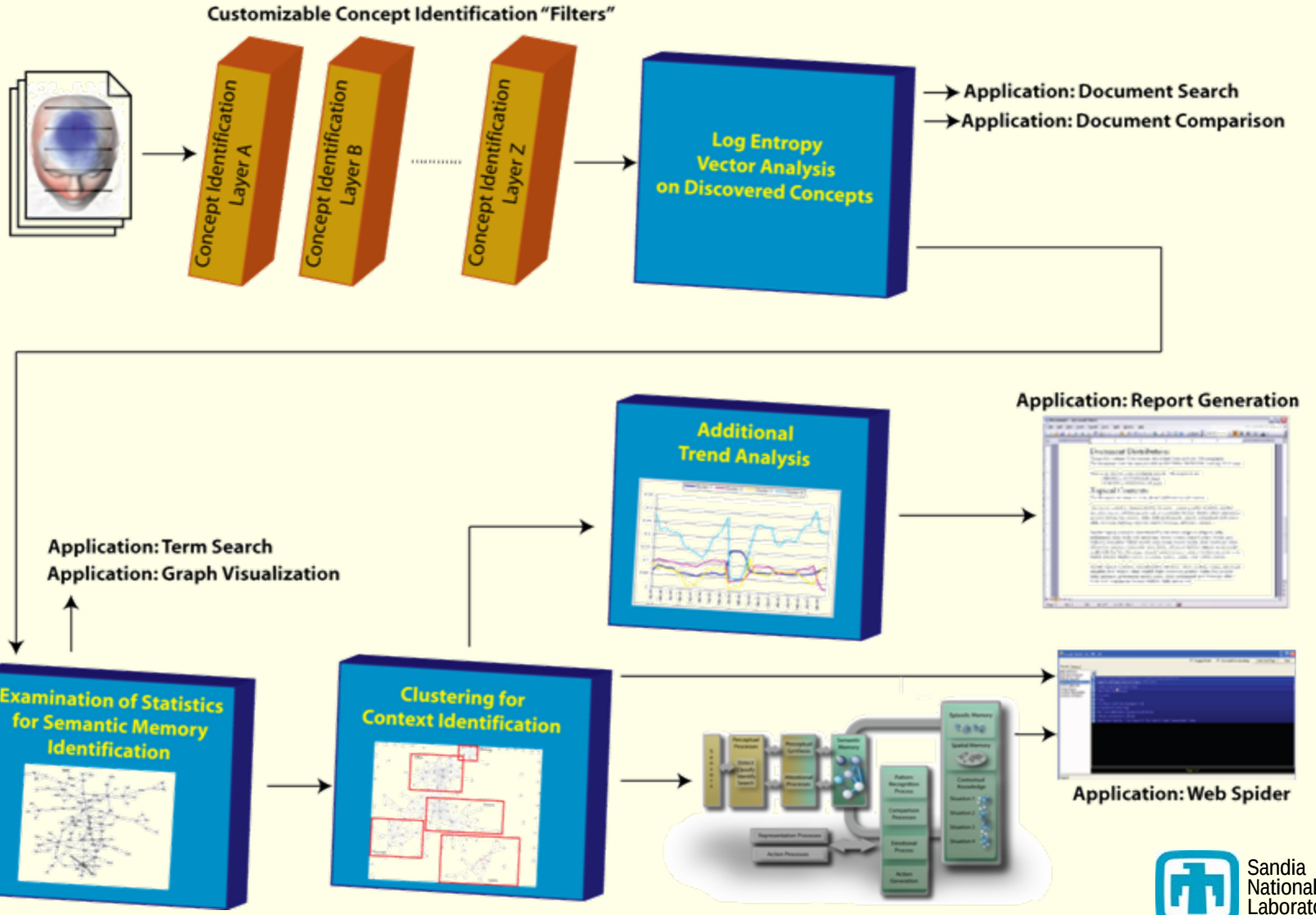
TCP/IP

Application Development

STANLEY Application Components (Advanced Capability)

STANLEY Library (Core Capability)

STANLEY Components



Cognitive Modeling through Text is not . . .

- Information Retrieval
  - Goal is not to retrieve information
  - The corpus itself is not what is of interest. It's what stands behind the corpus.
- Knowledge Extraction
  - Do not work on the assumption that the "knowledge" is something in the text itself.
  - Knowledge exists only in the interaction of the data and a cognitive system
- Data Mining
  - "The science of extracting useful information from large data sets or databases" (D. Hand, H. Mannila, P. Smyth: Principles of Data Mining. 2001)