Living with STANLEY:
Text Analysis for Cognition Research, Development, and Deployment

Travis Bauer

Cognitive and Exploratory Systems and Simulation
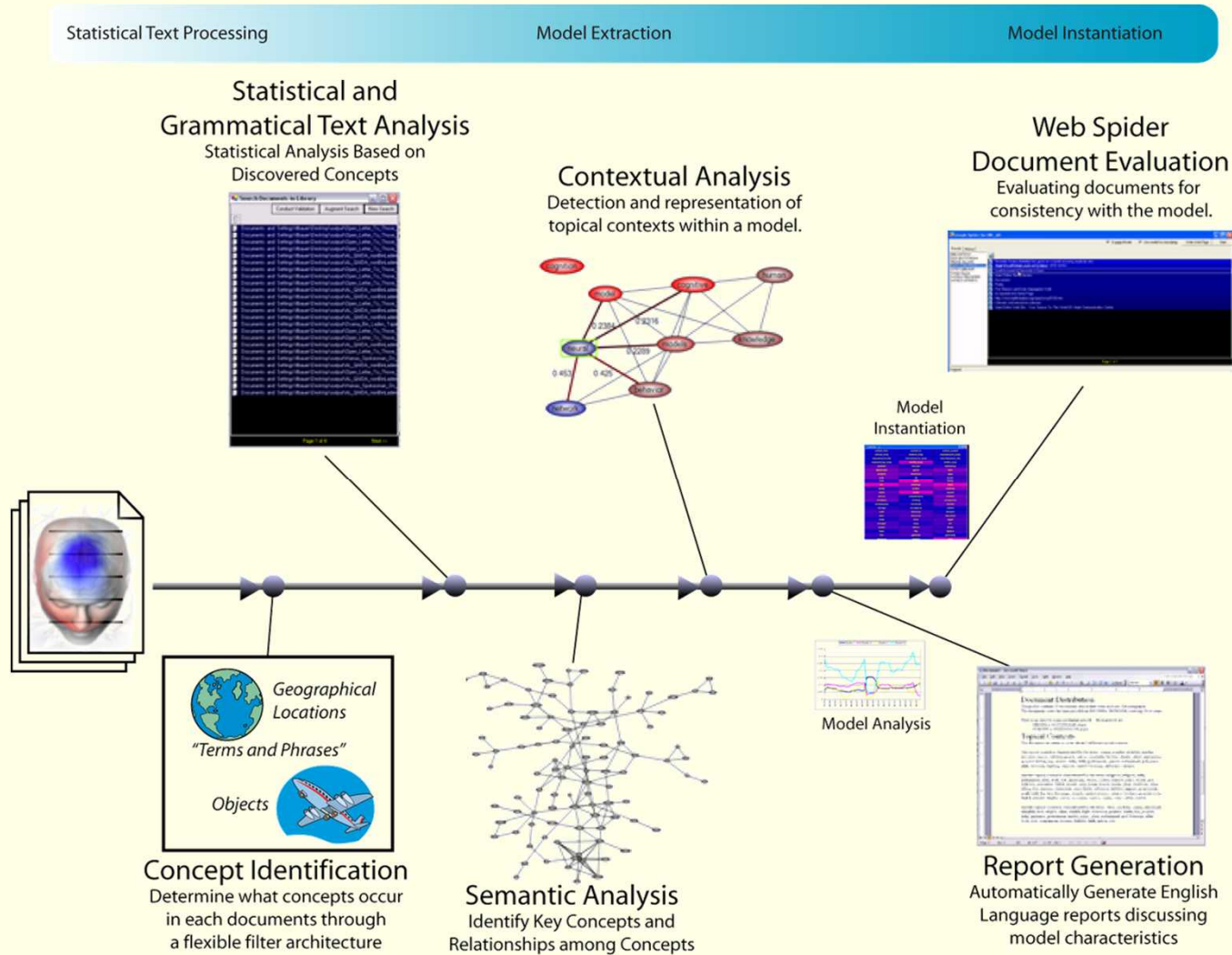
tlbauer@sandia.gov

284-8723

Sandia National Laboratories

Outline

- **What is STANLEY?**
- **Where did STANLEY come from?**
- **What applications are built on it?**
- **Where is it going?**

What is STANLEY?
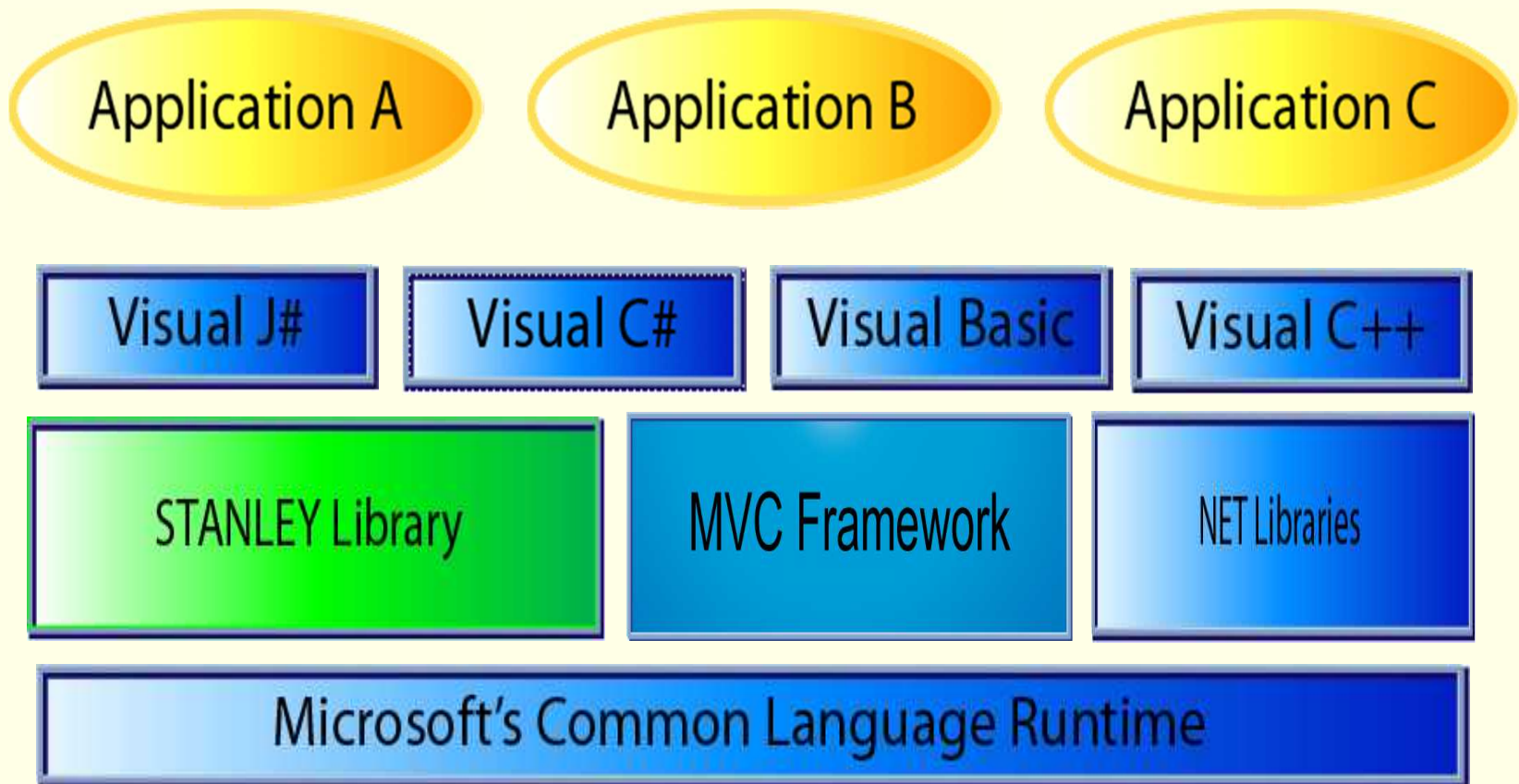
STANLEY is a statistical text analysis library designed for
# Automatically characterizing a person from text the produce or access

Statistical Text Processing · Model Extraction · Model Instantiation

Statistical and Grammatical Text Analysis
Statistical Analysis Based on Discovered Concepts

Contextual Analysis
Detection and representation of topical contexts within a model.

Web Spider Document Evaluation
Evaluating documents for consistency with the model.

Model Instantiation

Geographical Locations
"Terms and Phrases"
Objects

Concept Identification
Determine what concepts occur in each documents through a flexible filter architecture

Semantic Analysis
Identify Key Concepts and Relationships among Concepts

Model Analysis

Report Generation
Automatically Generate English Language reports discussing model characteristics

Sandia National Laboratories

STANLEY is built on the .NET platform

Where did STANLEY come from?

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.
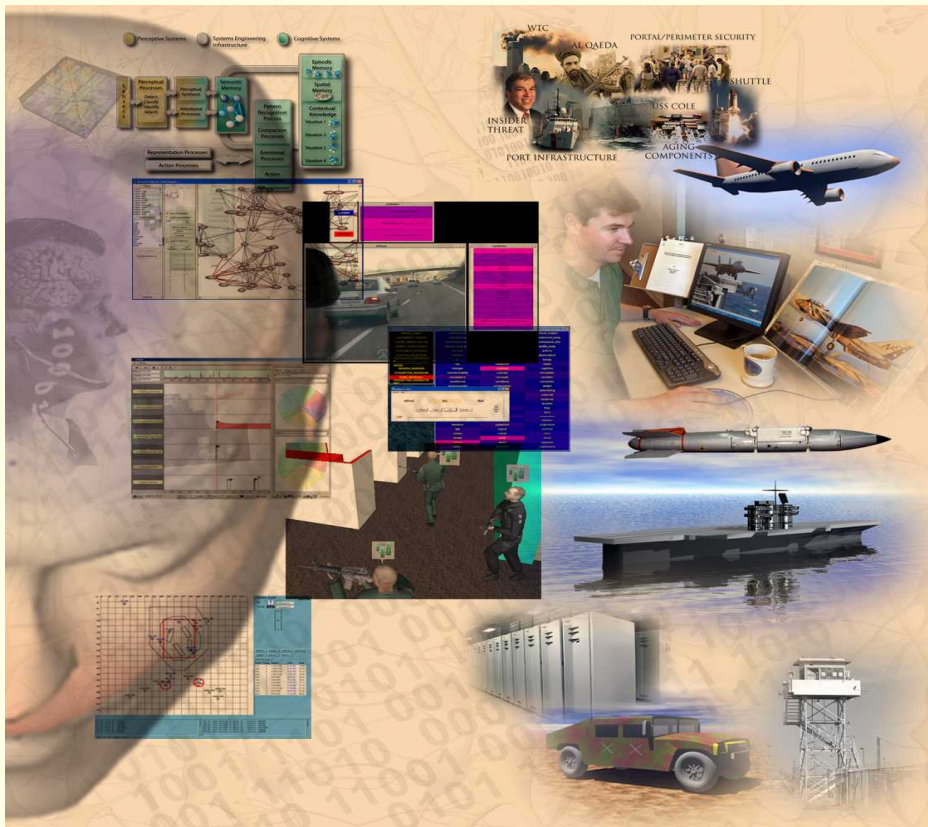
- **Program start 1999.**

- **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.**

Sandia National Laboratories

AugCog Applications and Prototypes
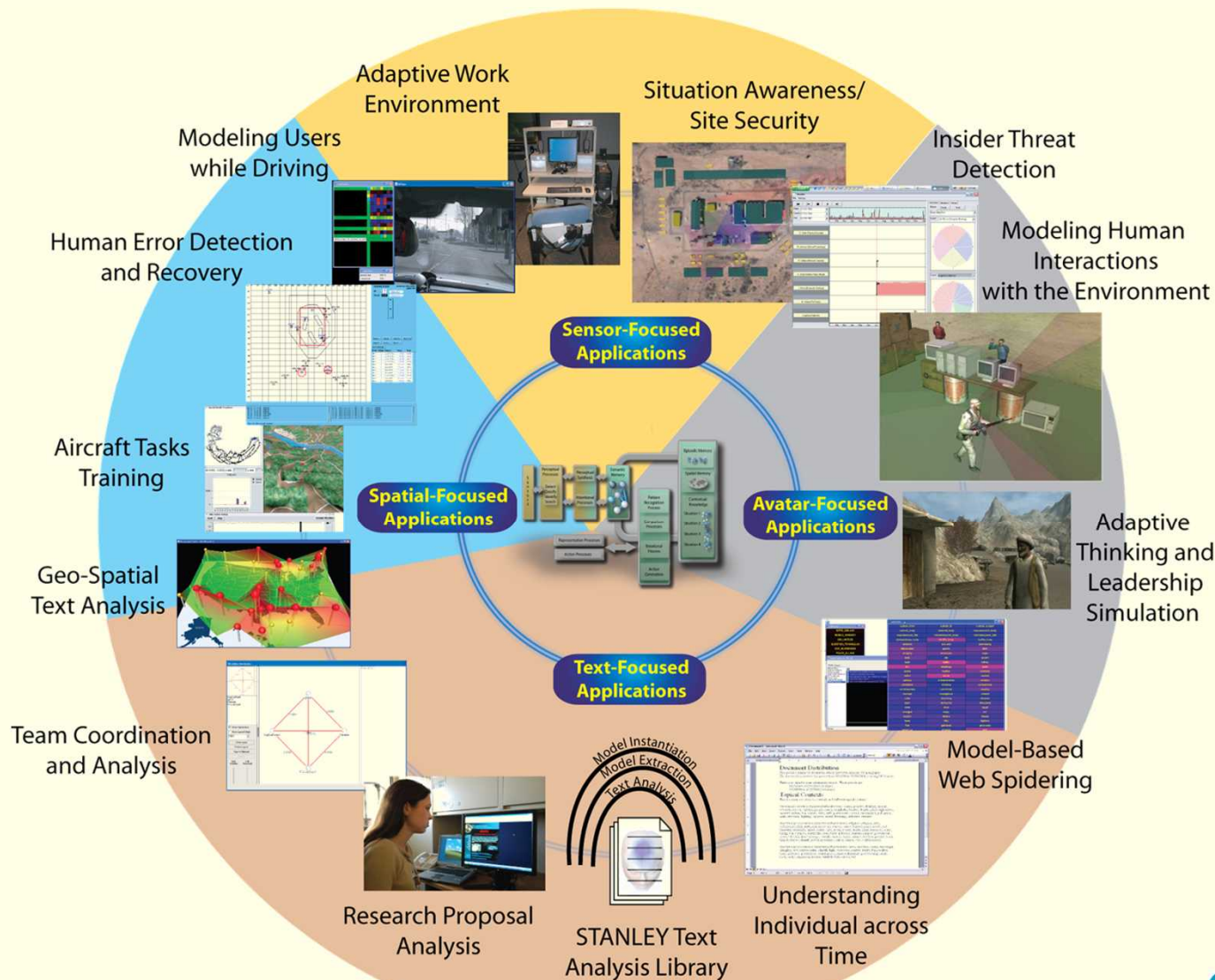
STANLEY Development Timeline

- **Original Problem: Automated creation of models for the cognitive model framework**
- **Original Task (May 2003): Collect documents people access or produce**
  - **Targeted Windows-based computers**
- **Task extended to sample demo application called "Cognitive Aide" designed to analyze the individual using the software**
- **Eventually, most requests to use the software were to build models of other people or data sets**
  - **Separated out the analysis capability independently**
  - **Built the first version of the Analyst Aide**
- **With infusion of funding from new sources, built additional applications, and are improving STANLEY itself**

Sandia
National
Laboratories

What applications are built on STANLEY?

Sandia
National
Laboratories

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

Cognitive Systems Applying computer models of human cognition to create unique technology solutions
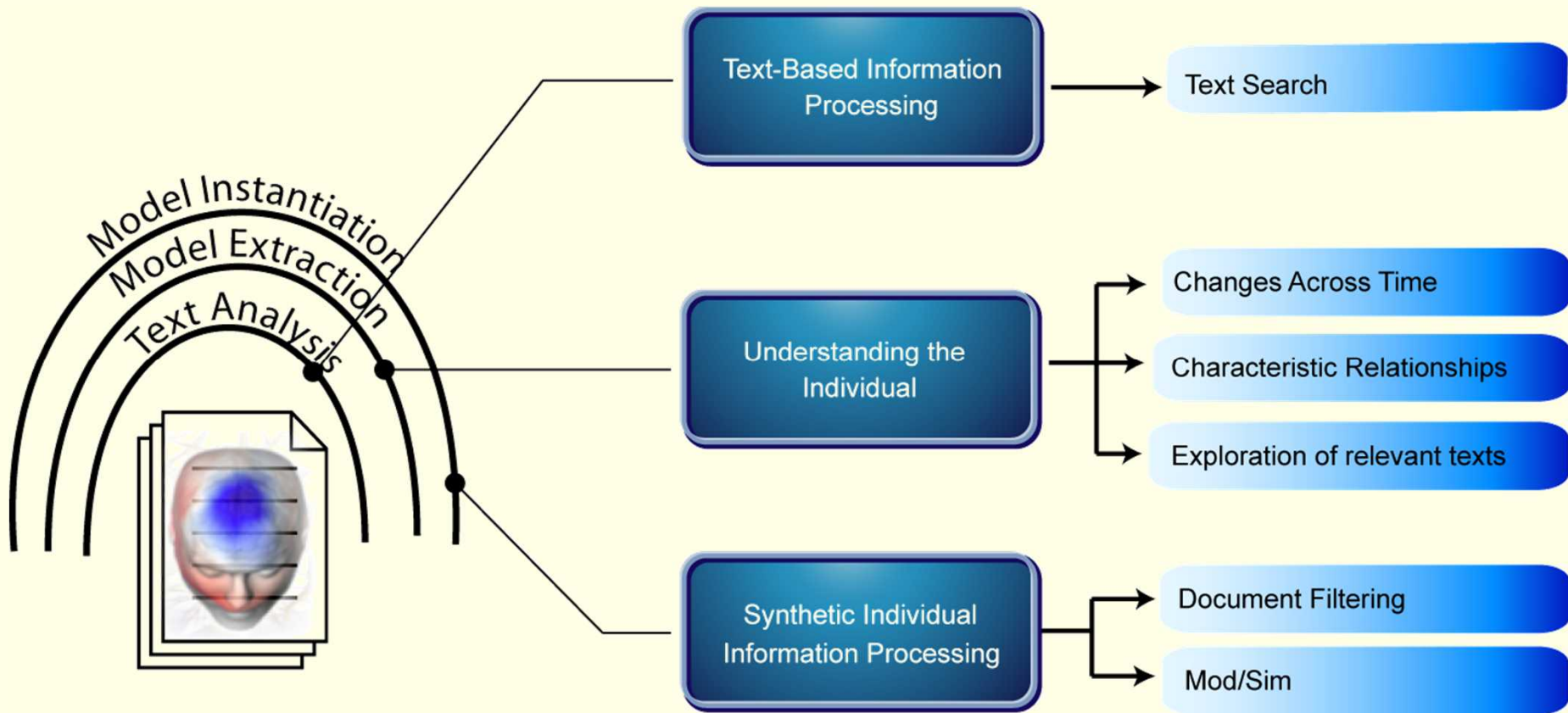
Problems We've Addressed with this Technology

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.**

Sandia National Laboratories

STANLEY Related Products

- **STANLEY**
- **Sandia Analyst Aide**
- **GeoViz / LDRD View**
- **Sounding Board**
- **ESC Investigator**
- **Data Warehouse Server/Client**
- **Ontology Helper**
- **2 Tools for the Yucca Mountain Project**

Sandia National Laboratories

Where is STANLEY going?

Current Development

- **Current development**
  - **12 people working on STANLEY related efforts directly through our organization**
    - **Unified development processes**
    - **Integrated team approach to working on multiple projects**
  - **Testing and hardening code for production use**
  - **Increasing collaborations with other Sandians**
  - **Rewriting Analyst Aide with the MVC Framework**
  - **Improving and growing the MVC Framework**
  - **Greater research focus**
- **Future Goals**
  - **JAVA – STANLEY 2.0 will be in Java and will be cross platform**
  - **Open source/Extendable Analyst Aide**
  - **Integrated, cross organization library**
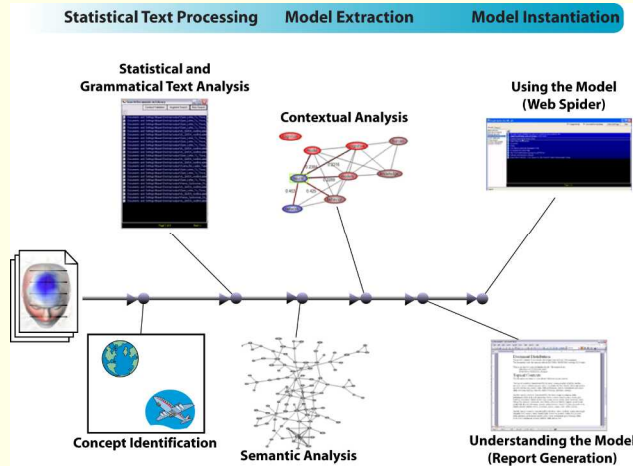 - **User's group to assist with use and collect feedback for future versions**

Sandia National Laboratories

Cognitive Systems Applying computer models of human cognition to create unique technology solutions

We want more collaborators!

- **Areas of interest**
  - **Visualization**
  - **Latent Semantic Analysis**
  - **Entity/Event Extration**
  - **Classification Algorithms**
- **Library/Applications available internally for testing**
  - **http://socrates.sandia.gov/twiki/bin/viewauth/CogSysPublic**
  - **Low traffic mailing list**

Sandia National Laboratories

Backup Slides

Cognitive Systems
Applying computer models of human cognition to create unique technology solutions

STANLEY Library

Statistical Text Processing   Model Extraction   Model Instantiation

Statistical and Grammatical Text Analysis

Contextual Analysis

Using the Model (Web Spider)

Concept Identification

Semantic Analysis

Understanding the Model (Report Generation)

The STANLEY library uses different techniques to create models.

What is it?

Text analysis library designed for building and utilizing models of individuals from text that they produce or access
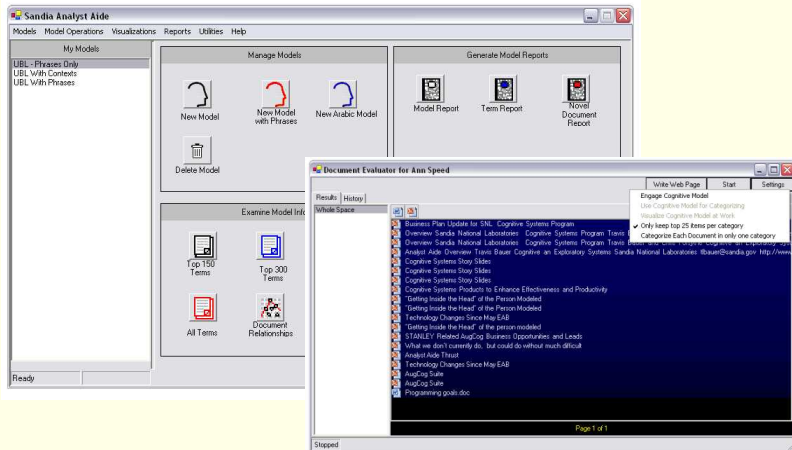
What are example applications?

- **Any application where one wants to integrate Sandia's cognition text analysis capability into some new application.**
- **Finding new relevant documents**
- **Understanding how a person or source has changed across time.**

What are the main features?

- **Reusable library that programmers can implements into new applications.**
- **Flexible implementation of text analysis algorithms**
- **Identification of key concepts, relationships, and contexts from text documents.**
- **Web spider for evaluating new information based on a previous analysis.**
- **Generation of reports in MS Word describing key characteristics of an individual based on the text analysis.**

Sandia National Laboratories

Sandia Analyst Aide

The Analyst Aide exposes most of STANLEY's functionality

What is it?

Application for building models of individuals from text they produce as well as understanding and utilizing those models.
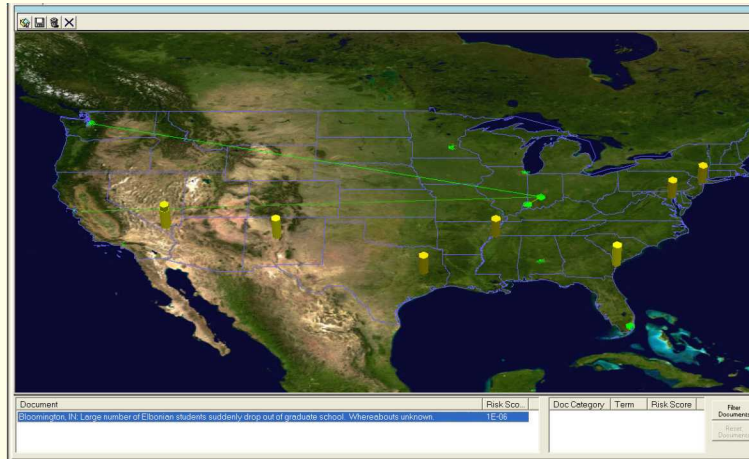
What are example applications?

- **Understanding the "big picture" of what the key concepts and relationships are for a person.**
- **Look at how a particular term or group of terms has become more or less important across time.**
- **Find documents that a person is likely to find of value.**
- **Query a model of an individual for related terms and documents.**

What are the main features?

- **Convenient interface to most of the features in the STANLEY library.**
- **Manage multiple models at one time**
- **Track how a person changes across time.**
- **Use a model to find new information on the web or evaluate new documents.**
- **Create graphical visualizations of the relationships among models.**

Sandia National Laboratories

GeoViz

Documents can be plotted geographically

What is it?

Server version of STANLEY with the capability of displaying documents geographically.

What are example applications?

- **Letting a third party application provide data or query STANLEY across a network.**
- **Letting a user see how geographical locations that may be related based on available documents.**

What are the main features?

- **Server interface lets other applications add in new documents and query for information through a network interface using XML**
- **Identify the geographical locations mentioned in a document.**
- **Plot documents on a map based on the geographical locations listed in them.**
- **Shows links among documents graphically on the map.**

Sandia National Laboratories

Sounding Board

Sounding Board

| About | Advanced Interface |

| Matching Models | Key Terms | Matching Paragraphs |

Aristotle   Descartes

Babbage   John Quin...

nature
power
declaration
delegating
sovereignty
dissolve
tyranny
stopped
spontaneous
signers
separately
precisely
degenerated

"file:Orations.txt"
the omnipotence of the God of battles. Union! Union! was the instinctive and simultaneous cry throughout the land. Their Congress, assembled at Philadelphia, once--twice--had petitioned the king; had remonstrated to Parliament; had addressed the people of Britain, for the rights of Englishmen-- in vain. Fleets and armies, the blood of Lexington, and the fires of

"file:C:\Documents \Demos\STANLEY \Orations.txt"
*The Project G Quincy Adams Adams This E Etext of the m Anniversary of

(Categories not

User can query multiple expert models at one time.

What is it?

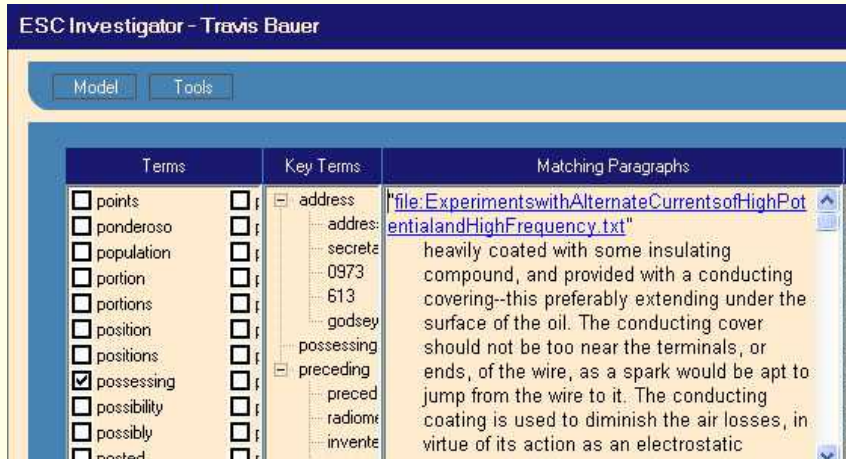**An application that lets you query and explore multiple models at one time.**

What are example applications?

- **Helping the user determine what human expert most likely understands a particular situation.**
- **Helping a user to find new related concepts they may not have previously known about.**
- **Helping the user to find texts showing how an expert uses particular terms.**

What are the main features?

- **Free text query allows the user to describe a situation.**
- **Multiple models read query and display themselves if they think it is in their area of expertise.**
- **User can find related terms for a particular model as well as texts used to create the model.**

ESC Investigator

ESC Investigator - Travis Bauer

Model   Tools

Terms	Key Terms	Matching Paragraphs
☐ points	⊟ address	"file:ExperimentswithAlternateCurrentsofHighPot
☐ ponderoso	addres:	entialandHighFrequency.txt"
☐ population	secreta	heavily coated with some insulating
☐ portion	0973	compound, and provided with a conducting
☐ portions	613	covering--this preferably extending under the
☐ position	godsey	surface of the oil. The conducting cover
☐ positions	possessing	should not be too near the terminals, or
☑ possessing	⊟ preceding	ends, of the wire, as a spark would be apt to
☐ possibility	preced	jump from the wire to it. The conducting
☐ possibly	radiome	coating is used to diminish the air losses, in
☐ posted	invente	virtue of its action as an electrostatic

ESC Investigator provides flexible exploration of document sets

What is it?

An application for exploring a particular set of documents
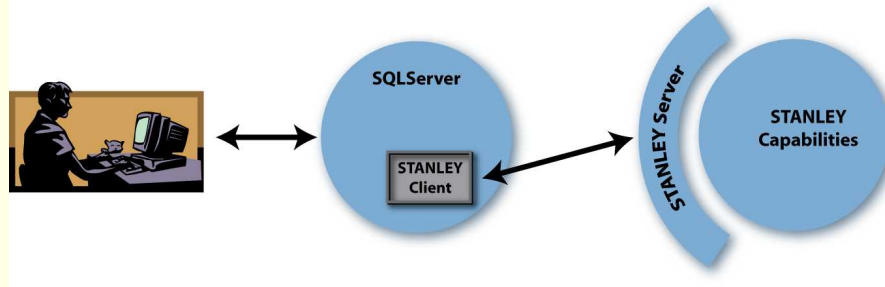
What are example applications?

- Helping a user to find new related concepts they may not have previously known about.
- Helping the user to find texts showing how an expert uses particular terms.

What are the main features?

- Free text query allows the user to describe a situation.
- Check box listing all terms so the user can select terms of interest.
- User can find related terms for a particular model as well as texts used to create the model.

Sandia National Laboratories

Data Warehouse Server and Client

What is it?

An interface to STANLEY that lets it's features be accessible directly from a SQL server database

The user accesses STANLEY analysis through an SQL server
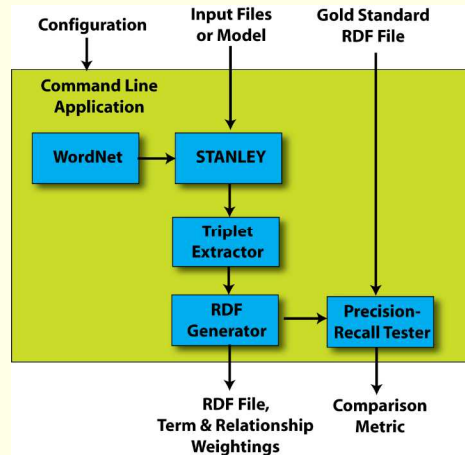
What are example applications?

- **When a user wants to create a model of an individual, but the individual's documents are located in a database.**
- **When you want to deploy the software to an organization, but don't want to install the software on every individual's desktop.**

What are the main features?

- **Complete integration with SQL server – STANLEY features become available as SQL functions.**
- **Flexible server architecture lets the analysis be done on any machine, not just the server.**
- **Querying for documents and key terms.**
- **Querying for changes across time.**

Sandia National Laboratories

Ontology Helper

A multi-stage process is used to create seed ontologies

What is it?

A command line application that finds key terms and relationships for an ontology from a corpus of text.

What are example applications?

- **When the user wants to create an ontology from a large corpus and doesn't know where to begin. This application can create a "seed" ontology which the user can subsequently edit.**

What are the main features?

- **Extraction of triplets (subject/verb/object) from text.**
- **Identification of key terms and phrases.**
- **Export to RDF format for compatibility with other ontology tools.**

Sandia National Laboratories