

**The DHS  
University Affiliate Centers  
to the  
Institute for Discrete Sciences  
(IDS-UACs)**

**DyDAn  
MIAS  
CERATOPS  
CKID**

**In collaboration with  
Lawrence Livermore  
National Lab**

Fred Roberts, Rutgers University

# What is IDS?

**The Institute for Discrete Sciences is a collaboration among DHS, the National Laboratories and four University Affiliate Centers of Excellence, led by Lawrence Livermore National Laboratory**



IDS focuses on selected topics in computer science, data sciences, machine learning, and discrete mathematics, collectively labeled "discrete sciences"

# IDS is a “Center of Centers”

RUTGERS

**DyDAn:** The Center for Dynamic Data Analysis



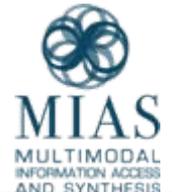
**CKID:** The Center for Knowledge Integration and Discovery



**CERATOPS:** Center for Extraction and Summarization of Events and Opinions in Text



**MIAS:** Multimodal Information Access and Synthesis



Collaborating with **Lawrence Livermore National Lab**

# The IDS Partner Network



(coordinator)



# IDS mission

- Homeland Security critically depends on ability to **draw inferences from massive amounts of unstructured data**
- Mission:
  1. **Develop technologies** that provide the machinery to derive knowledge from data
  2. **Educate students** in Homeland Security technology

*The fact is that systematic intelligence lies at the heart of everything that we do. Understanding the enemy's intent and capabilities affects how we operate at our borders, how we assess risk in protecting infrastructure, how we discern the kind of threats for which we must be prepared to respond.*

Michael Chertoff

*Knowledge will forever govern ignorance; and a people who mean to be their own governors must arm themselves with the power which knowledge gives.*

James Madison

# IDS application areas

- IDS methods are applicable to a wide variety of homeland security applications.
  - Intelligence Analysis of Text
  - Disease Event Detection
  - Port of Entry Inspection

A photograph of a man and a woman wearing face masks. The man is holding a mobile phone to his ear. They are standing in front of a banner that says "you need help" and "your information at 区长信箱".

**SARS**

来日決定！**JAPANツアー2003**

札幌ドーム・仙台市民会館・東京ドーム2DAYS  
名古屋ドーム・大阪ドーム・福岡ドーム

お問い合わせは「まにあつくす on the WEB.」まで

**NEW ALBUM 「Panic」 NOW on SALE**



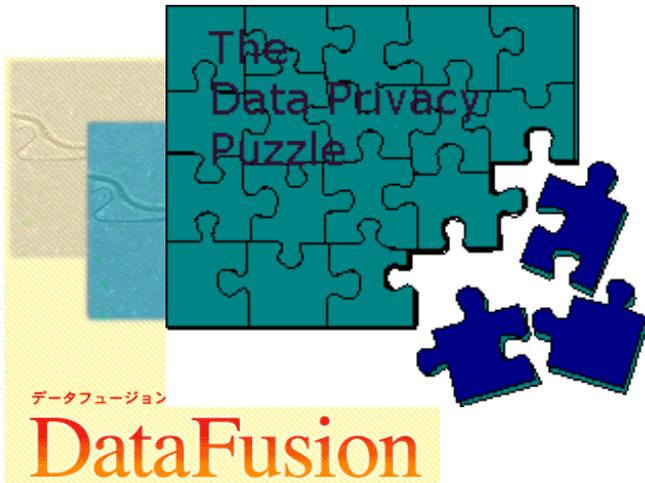
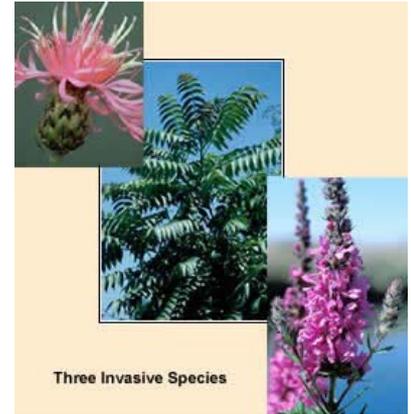
# IDS application areas

- IDS methods are applicable to a wide variety of homeland security applications.
  - **Author Identification**
  - **Response to Natural Disasters**
  - **Bioterrorism**
  - **Sensor Location**



# IDS application areas

- IDS methods are applicable to a wide variety of homeland security applications.
  - **Protection Against Invasive Species**
  - **Customs and Border Protection**
  - **Privacy-Preserving Data Sharing**



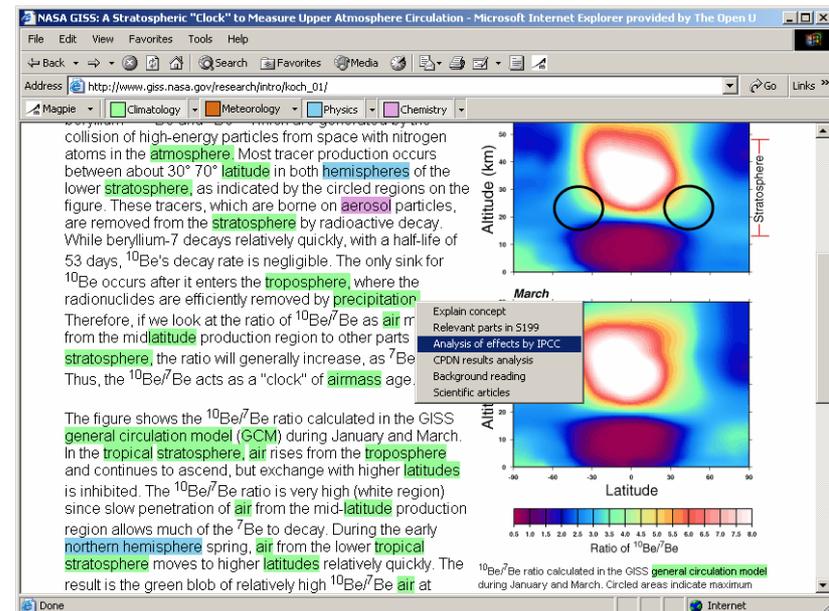
# IDS research goals

IDS seeks to develop fundamental theories, algorithms, and tools that enable us to gain knowledge from data

- **Problem 1: Extracting pertinent information from various media**

- New capabilities in information identification, extraction, storage, and access across media

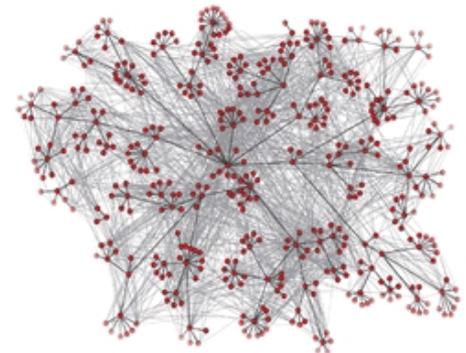
- E.g., robust techniques for extracting, summarizing and tracking information about events from unstructured text



# IDS research goals

IDS seeks to develop fundamental theories, algorithms, and tools that enable us to gain knowledge from data

- **Problem 2: Fusing and storing this information**
  - Computationally efficient methods for representing and fusing information, while preserving privacy
  - E.g., representing and storing info in semantic multigraphs



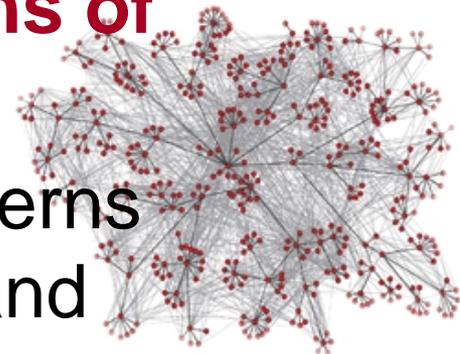
# IDS research goals

IDS seeks to develop fundamental theories, algorithms, and tools that enable us to gain knowledge from data

- **Problem 3: Finding trends and patterns of interest**

- Novel technologies for identifying patterns and relationships in massive graphs and datasets that change rapidly

- E.g., trainable learning algorithms to discover and extract events such as infectious disease outbreaks



pandemic flu

# IDS educational mission

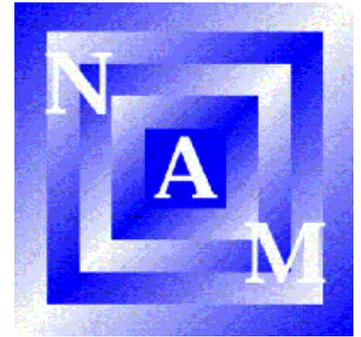
## Broad Goals:

- Establish pioneering educational programs that nurture the homeland security workforce of the future
- Develop curricular materials for all levels
  - We are developing materials targeting high school students through college and university faculty



# IDS educational mission

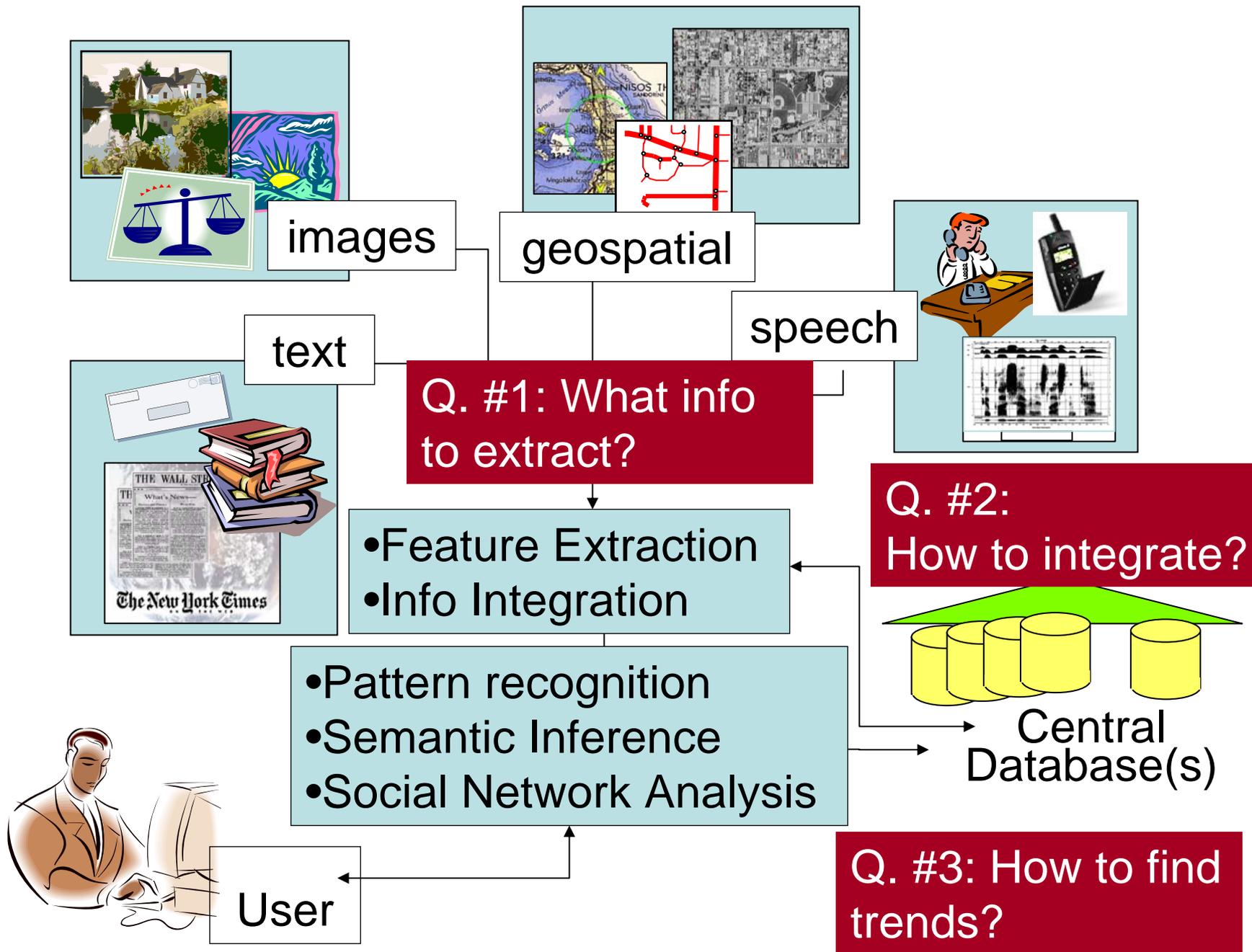
- Encourage participation of traditionally under-represented groups
  - We are committed to broadening participation
  - We are working with the National Association of Mathematicians – dedicated to enhancing the involvement of minorities
  - The United Negro College Fund Special Programs Corporation
  - A variety of programs for women
  - Many minority-serving partner institutions



# IDS educational mission

- **Sampling of planned events:**
  - Semantic Graphs workshop (joint among 4 IDS UACs)
  - July: 8-week summer school (UIUC)
  - August: Data Analysis in Law Enforcement and Homeland Security (Rutgers)
  - May: Mathematics and Homeland Security for High School Teachers (Rutgers)
  - October: Data Privacy workshop (Rutgers)
  - Summer Internships (all 4 UACs)





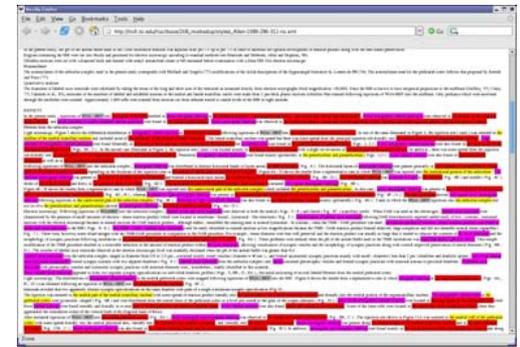
# Information Extraction

- **Information from Text:**

- Develop learning methods to perform extractions automatically
- Identify basic information like **entities & events**
- Determine if a sample is from an entity, as in author identification

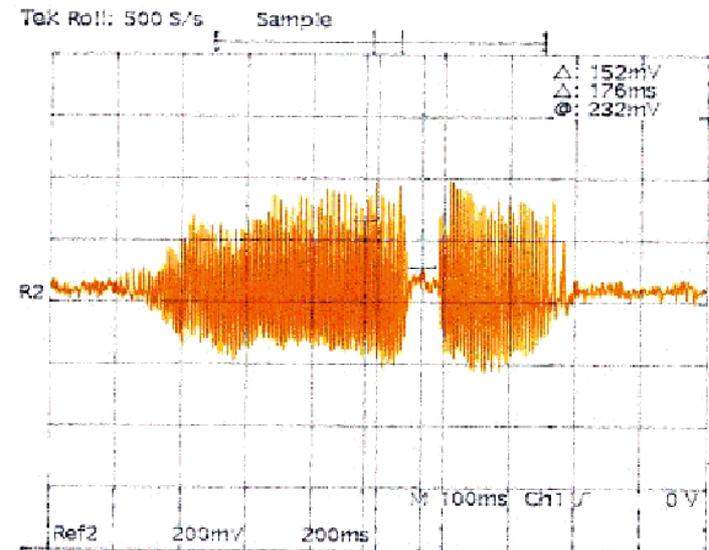
- **Information from Geospatial Data:**

- Develop methods to overlay and connect **maps, satellite images, GPS coordinates, spatial diagrams**, etc.
- Link this to addresses, names, disease syndromes, locations of supplies for natural disasters, etc.; connect to databases of various kinds



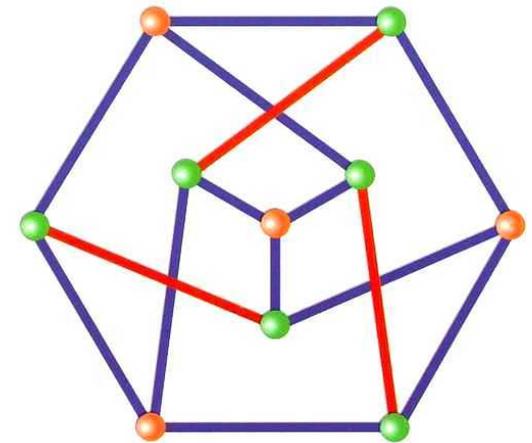
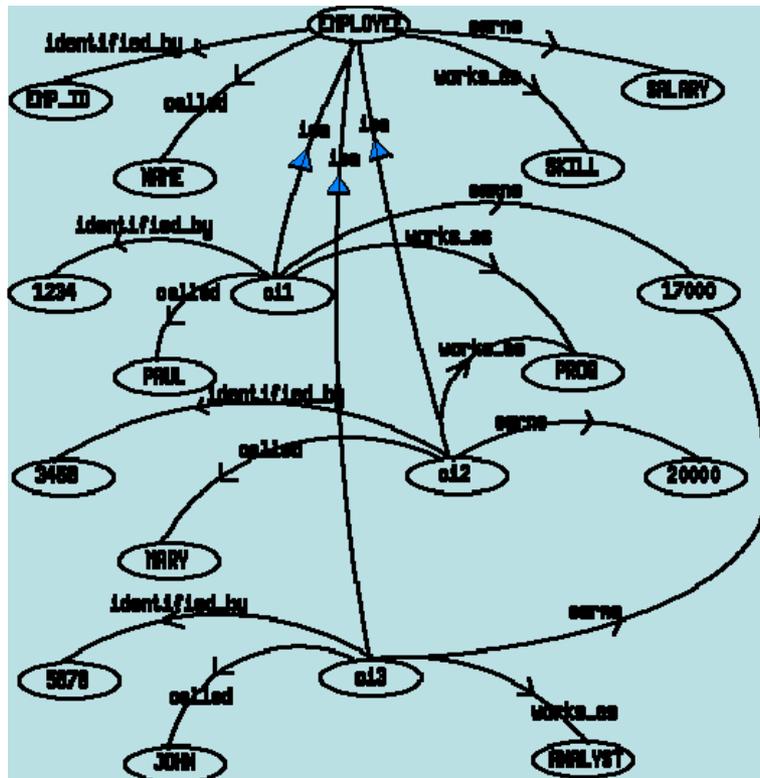
# Information Extraction

- **Information from Images & Speech:**
  - Images: From basic component features (colors, textures, etc.) to **identification of people** or **identification of invasive species** or **identification of dangerous cargo**
  - Speech: **Speech-to-text** conversion and **identification of emotion, stress, etc.**



# Representing and Fusing Information

- **Multigraphs:** Represent source information that is cross-linked multiple times by various relations
  - Fuse, query, navigate, and browse
  - Entity correlation, identity discovery



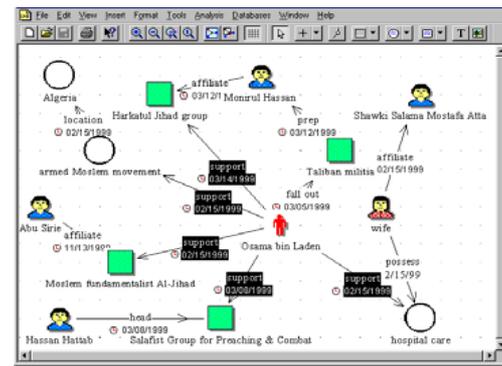
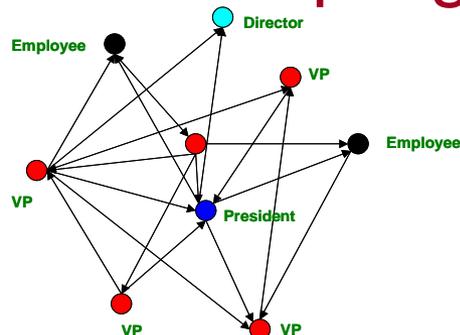
# Representing and Fusing Information

- **Data management:** Large databases
  - Anonymization, storage
  - Assure appropriate levels of privacy
  - Massive **amounts of data, distributed, continuously changing:** use Bayesian regression and filtering
  - Interpreting patterns of sensor data, as in bioterrorism sensors or critical infrastructure protection
- **Adding semantics:** Deploy large general ontologies over databases
  - Develop methods for **cross-metadata and cross-ontology matching**



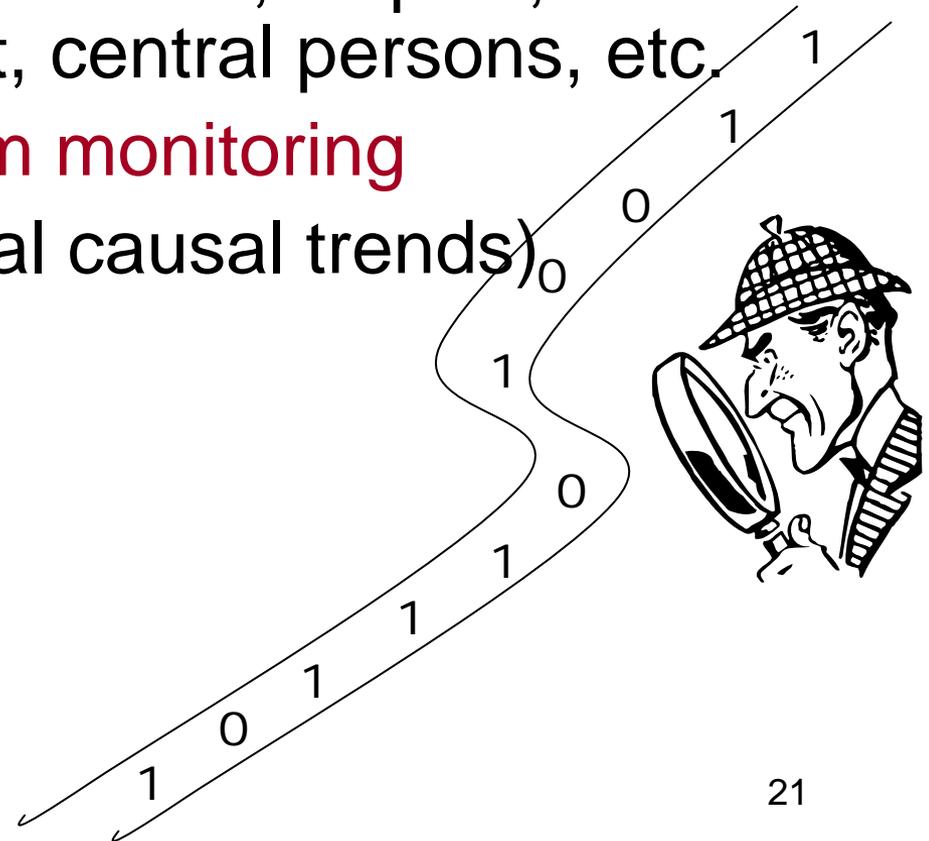
# Retrieving and Accessing Info

- **Textual information retrieval:**
  - Methods to **enhance information search** over the web and the 'deep' web
- **Database retrieval and optimization:**
  - Methods for handling very large amounts of changing data
  - **Similarity search and dimension reduction**
  - **Inference-based search** and knowledge base management
  - **Optimal sampling** from massive datasets



# Pattern Detection

- **Detect correlations and patterns over large (multi)graphs:**
  - **Social Network Analysis** of email networks; detect hidden social structures, cliques, communities of interest, central persons, etc.
  - Continuous **data stream monitoring**
  - **Text entailment** (general causal trends)
  - Analysis of **blogs**



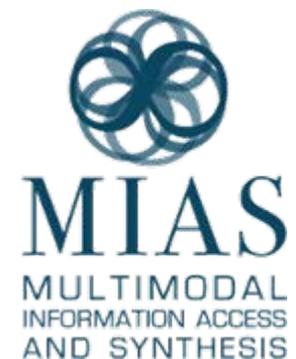
# Pattern Detection

- **Statistical & information-theoretic techniques for pattern and trend discovery over multi-media info:**
  - Apply to **data streams in multidimensional space**
  - Apply to **time-varying multigraphs**
- **Inference-based pattern detection and explanation:**
  - Look for pattern departure as in port of entry inspection
  - **Rapid detection of credit card fraud, network attacks, cyber attacks.**



# Concluding Remarks

- ***Discrete science is a “cross-cutting” capability that can provide support to each of the DHS divisions***
- We want to partner with other Centers, the DHS divisions, and the broader homeland security enterprise



The IDS Centers

# Please Contact Us:

- **DyDAn**: The Center for Dynamic Data Analysis
  - Fred Roberts ([froberts@dimacs.rutgers.edu](mailto:froberts@dimacs.rutgers.edu))
- **CKID**: The Center for Knowledge Integration and Discovery
  - Ed Hovy ([hovy@isi.edu](mailto:hovy@isi.edu))
- **MIAS**: Multimodal Information Access and Synthesis
  - Dan Roth ([danr@cs.uiuc.edu](mailto:danr@cs.uiuc.edu))
- **CERATOPS**: Center for Extraction and Summarization of Events and Opinions in Text
  - Jan Wiebe ([wiebe@cs.pitt.edu](mailto:wiebe@cs.pitt.edu))