Knowledge Discovery & Dissemination

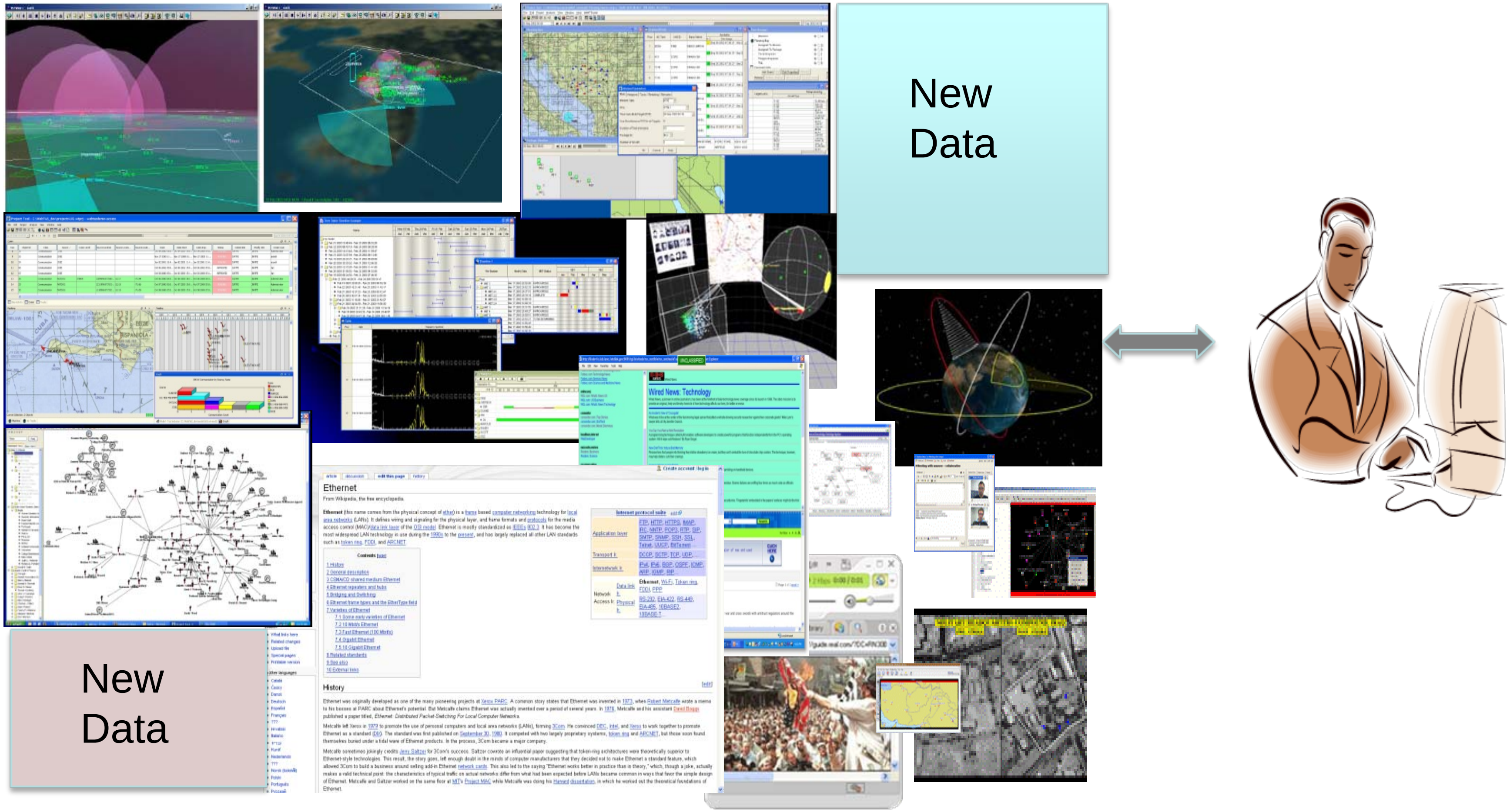
Enabling Analysts To Quickly Produce Actionable Intelligence From Multiple Sources of Information

Dr. Arthur Becker

ODNI/IARPA/Incisive Analysis

Knowledge Discovery and Dissemination

Enabling Analysts to Produce Actionable Intelligence from Multiple Sources



Knowledge Discovery and Dissemination

Tasks

Data Alignment Research toward automating the semantic alignment of multiple data sets including new and unfamiliar data sets

Advanced Analytic Research to develop powerful analytic tools that work across multiple disparate data sets similar to tools that work within a single data set

Prototype Development that implements research algorithms and can be tested against IC problems

Evaluation

Measure performers research through the performance of their prototypes against real IC problems, real IC data and used by analysts

What We Aren't Doing

-- Scalability research

-- Media processing

-- User interface research and user studies

-- Foreign language processing

Alignment Problem

Data bases created by others are organized and use terminology to support their needs. Terms could be different or could assign different meanings to the same term

The concept of a location could be labeled as” Address, Place, Location, Locality, Point and other ways

Meaning: Address, Lat-Long, Grid coordinate, District (police, school, political corporate…), Region County, Neighborhood

Even with specific meanings, the expression could be very different:

“RT 5 Box 2340” or “12345 Main Street” or “School House Hill” are legitimate postal addresses for the same place

Some of these are 1-1 mappings more often they are not, for example:

Professional < ----- > Dentist

Alignment Technical Approaches

Folksonomics: Analysts and Subject Matter Experts providing guidance

Data Driven: Facts and relationships extracted from raw text

Context & Usage: Data model extracted from probability distribution; function of terms used

Top Down: High level ontology and domain ontologies

Hierarch of Ontologies: Combine high level and multiple domain specific ontologies

Solutions generally will combine multiple techniques

Advanced Analytic Research

Most advanced analytic tools are tailored to a specific type of analysis and/or fixed data types

KDD research is focused on extending these techniques to situations that are more general

Techniques Proposed	Capability Provided (If It Works)
Generalized search by example	Given a number of examples, find the common thread and find similar items
Generalized social networks	Use multiple types of relationships over time to understand a network of people
Context of loose term	Put useful definitions to terms like “near” or “similar”
New mathematics for categorization (i.e. replace LSI and LDA with beta processes and new variants of LDA)	Find hidden relationships not explicitly in the data (e.g. bombing and financing)

Evaluation Metrics

- Metrics are based on the performance of the prototype over the analytic test range. Prototypes will be measured in terms of how accurately, completely and quickly they perform tasks
- Alignment time will be restricted and reduced for each later cycle
- All tests will be objective, repeatable and statistically valid
- Statistical validity will be accomplished by use of sufficient number of analysts
- Test platforms will be instrumented to collect detailed performance data

Research Teams

Applied Communication Sciences	BAE Systems	CUBRC	SRI International
Rutgers University Intuidex, Inc. University of Illinois	Brown University, Carnegie-Mellon University University of Massachusetts, Amherst Lymba Corporation	State University of New York-Buffalo Intelligent Software Solutions General Dynamics Securboration	University of Washington University of California-ISI Stanford University New York University (NYU) Carnegie-Mellon University Oculus, Inc.