New Directions in Digital Government: Knowledge Discovery and Dissemination

Gary W. Strong

Computer and Information Science and Engineering

National Science Foundation

NSF has a broad, cross-cutting research agenda that includes national security.

Information Technology Research (ITR) Homepage - Microsoft Internet Explorer

File   Edit   View   Favorites   Tools   Help

itr
Information Technology Research

Software & Hardware Systems

Augmenting Individuals & Transforming Society

Scientific Frontiers & Information Technology

Program Announcement NSF 01-149 FY 2002

FY 2001 Awards
FY 2000 Awards

News & Media

Participating NSF Directorates & Programs

FAQs

ITR: Division and Program Selection

NSF Interest in ITR Projects Related to National Security

Home   About ITR   Site Map

The National Science Foundation
4201 Wilson Boulevard,
Arlington, Virginia 22230, USA

STIMULATE:
Speech, Text, Image, and MULtimedia Advanced Technology Effort

Program Solicitation
A JOINT INITIATIVE OF:

NATIONAL SCIENCE FOUNDATION COMPUTER AND INFORMATION SCIENCE AND ENGINEERING DIRECTORATE

NATIONAL SECURITY AGENCY OFFICE OF RESEARCH AND SIGINT TECHNOLOGY

OFFICE OF RESEARCH AND DEVELOPMENT CENTRAL INTELLIGENCE AGENCY

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY INFORMATION TECHNOLOGY OFFICE

DEADLINE: SEPTEMBER 1, 1996

last modified 11/01/2001

NSF's mission, set out in the NSF Act of 1950 (Public Law 810507) is:

To promote the progress of science; to advance the National health, prosperity, and welfare; to secure the National defense; and for other purposes.

Knowledge Discovery and Dissemination Working Group

- Part of interagency Intelligence Technology Integration Center of the Intelligence Community
- Started in early Summer 2001
- Now a funded program w/interagency steering committee
  - "Internal" prototype, working real problems, with real analysts working on real data
  - Leveraging existing research with 1-3 year payoff, with 12 efforts funded in FY02
  - Calls for new research with 5-7 year payoff, TBA
- KDD is a cross-cutting theme for other working groups/programs, with its topics:
  - Collaboration and Knowledge Sharing
  - Knowledge Representation
  - Knowledge Discovery in Massive Data Sets & Streams

KD-D Research Topics Funded in FY02

- Collaboration & Knowledge Sharing
  - Large-Scale Visual Interaction with Multi-Lingual Source Textual Data
  - Informedia Digital Library
  - Semantic Web Tools for Authoring & Using Analysis Results
- Knowledge Representation
  - Mining Multilingual Resources using Text Analytics
  - Statistical Semantic Parsing & Proposition Detection
  - "Talk Printing": Harnessing Speaking Habits for Speaker Recognition & Modeling
  - Infrastructure Support
- Knowledge Discovery in Massive Data Sets & Streams
  - Tools for Monitoring Online Information Sources
  - Patterns in Temporal Data, Unsupervised Algorithms
  - Monitoring Message Streams: Retrospective and Prospective Event Detection
  - Distributed Mining & Monitoring
  - Computational Statistics for Intelligence Analysis
  - Mining Frequent Structures in Massive & Complex Data

Unique KDD Issues

- Data Sharing
  - Solution will be a combination of policy and technology
  - One very-large database is illegal in the KDD context
  - Distributed queries are illegal to protect "unreasonable searches"
  - Distributed software agents are not trusted
  - Shared, auditable knowledge networks may fit legal definition of probable cause
- Data Mining
  - "Needle in the haystack" is an inappropriate metaphor
  - A very large jigsaw puzzle is more apt
  - Puzzle pieces are separately held and may not be mixed
  - Many data sets are streams, not bases

Interagency Cooperation on National Security Research

- Culture of academia changed on Sept. 11th
  - Researchers want to help
    - Need to know what the problems are
    - Workshops meet part of this need
- NSF is engaged
  - There is a large base of peer-reviewed grantees
    - A workshop-based process has been invented
    - It was used once in KDD and will be used in other domains
  - Funded projects are already relevant
    - Determine who gets to participate
    - Challenges of relevance have been met by simple searches over NSF award database

NSF - Funding - Award Data - Microsoft Internet Explorer

File   Edit   View   Favorites   Tools   Help

← Back ▾ → ▾ ⊗ ⬚ ⌂ | ⬚ Search ⬚ Favorites ⬚ History | ⬚ ▾ ⬚ ⬚ ▾ ⬚ ⬚

Address ⬚ http://www.nsf.gov/home/grants/grants_awards.htm   ▾ | ⬚ Go

About NSF   **Funding**   **Publications**   **News & Media**   **Search**   **Site Map**

Funding   Award Data

TOPIC:
General Info and Policies

▶ **Overview**
▶ **What's New**
▶ **Funding Opportunities**
▶ **Proposal Preparation**
▶ **Award Administration**
▶ **Award Data**
▶ **Budget Request**
▶ **Deadlines and Target Dates**
▶ **FastLane**
▶ **Overview of NSF Programs**

NSF provides several ways to research past (since 1989) and current awards:

Award Abstracts Database
Information about research projects that NSF has funded since 1989 can be found by searching the Award Abstracts database. The information includes abstracts that describe the research, and names of principal investigators and their institutions. The database includes both completed and in-process research. A Fielded Search is also available.

FastLane Award Searches The FastLane server offers several options, including:

- List of Recent Awards
- List of Awards by Program
- List of Awards by Institution
- List of Awards by State

Budget Internet Information System
This server provides summaries of award amounts by state, awardee institution, and NSF Directorate.

⬚   ⬚ Internet

The Big Challenge: Data

- Relevant datasets are needed to drive research
  - Credit card fraud, telecommunications fraud, insurance fraud, and identity theft are good domain examples
  - Those data contain protected personal information and are private assets to companies
- Artificial, simulated data is being proposed
  - Artificial data can be created in huge amounts (billions of transactions)
  - Scrubbing data is very expensive
  - Isn't "dirty" enough
- Data need to be coupled with evaluation, as in DARPA human language technology programs

Other Research Domains in Planning Stage

- Pattern Recognition/Characterization of People, Places and Things
- Spatio-Temporal Information Systems
- Biological Warfare Defense
- Sense-making/Augmented Cognition
- Signal Processing
- Global Change Detection
- Measurements and Signatures Intelligence
- Computational Social Systems
- Management of Knowledge-Intensive Organizations

Summary

- Agencies of Federal Government that didn't talk much are collaborating once again at a very deep level

- The Nation's best science is being called up to serve and wants to serve

- There are new processes being invented for addressing the needs

- The solutions are likely to be a combination of technology and policy