Knowledge Discovery and Dissemination Program (KDD)

Focus, Process, and Instructions for 2006

November 2005

This publication was produced by the KDD Program Staff as an aid to KDD Researchers, Project Monitors, and Members of the KDD Working Group. Questions and comments regarding this publication should be submitted to the KDD Program Staff through dlricha@afterlife.ncsc.mil

Knowledge Discovery and Dissemination Program (KDD)

Focus, Process, and Instructions for 2006

Introduction

This document provides background information and describes the process used by The Knowledge Discovery and Dissemination Program (KDD) in requesting, selecting, funding and testing research activities. Those who plan to submit white papers and/or research proposals for KDD funding in FY 2006 should submit their proposals in accordance with the instructions provided in this document.

Background

The Knowledge Discovery and Dissemination Program (KDD) is a research program sponsored by the Intelligence Community (IC) and managed within the Intelligence Technology Innovation Center (ITIC). The objective of the KDD program is to improve the quality and timeliness of intelligence analysis by creating effective methods to merge multiple sources of information and cooperatively analyze the information through multi-agency analytic teams. The KDD Program supports activities that help coordinate Knowledge Discovery research activities across the Intelligence Community. The KDD program funds research that is conducted at government agencies, laboratories, and academic institutions. The results of KDD research are provided to the KDD Program Manager and reports about the research are published in the Journal of Intelligence Community Research and Development (JICRD).

KDD Vision

Virtual, horizontal integration of intelligence analysis across multiple agencies covering diverse disciplines with access to multiple datasets using a simple rules-based tool to

collaborate, discover knowledge, share results, and disseminate all-source intelligence information to high priority customers.

Program Management

The KDD Program Manager uses a working group consisting of technical personnel and analysts from across the Intelligence community and its partners to maintain oversight of the program and coordinate KDD activities with the various agencies within the community. Representatives from CIA, DIA, DOD, NSA, NRO, NGA, DARPA, ARDA, DHS, Mitre, and NSF are members of the KDD Working Group. In addition to the KDD Working Group, the KDD Program Manager enlists volunteers as project monitors for each research project. The project monitors conduct direct technical interaction with the Principal Investigators. These interactions focus the substance of the research project and include attention to:

- Objectively assessing the value of the research effort from a technical and application point of view,
- Delivery of project results (e.g. software, algorithms, etc.) to the Chairman of the KDD Working Group on a CD,
- Submission of quarterly and an annual status report to the Chairman of the KDD Working Group on the research projects,

- Reporting the results of meetings with KDD researchers in a brief informal note to the Chairman, KDD Working Group,
- Ensuring that researchers fulfill the commitment to submit at least one article for publication in the Journal of Intelligence Community Research and Development (JICRD), and
- Confirming receipt of funds by the researchers and keeping abreast of project status.

Evolving KDD Program Focus

After surveying hundreds of programs and projects across the community, academia and industry, the KDD working group identified four areas that would have a significant impact and are appropriately pursued through an IC program. These are:

- Distributed Data Mining of multiple data sets and streams across different agencies
- Representing Target Knowledge for human and machine interpretation
- Sharing data in a way that coordinates constructively with law and policy
- Developing effective collaboration between analysts from a human factors perspective

Each of the four areas listed above expose difficult research problems as well as opportunities for collaboration with the larger research community. Pure research, by itself will not satisfy the overall objectives of the program. Solutions to IC problems will require integrating multiple components from each of the four areas into an end-to-end functionality. Therefore, the KDD program includes a test bed using real intelligence data and professional experienced analysts to solve critical and ongoing intelligence problems. The KDD program has developed an operational test bed, BLACKBOOK, as an evolving operational prototype that tests research deliverables. The testing process retains only the "best of breed" research solutions that are subsequently integrated fully in the following version of the prototype. BLACKBOOK serves to evaluate research products, refine needs and fully coordinate the research to operational reality and requirements.

During 2006 and beyond, the KDD Program will continue to address these four areas with specific focus on addressing:

- Weaknesses of evidence chaining including development of tools that allow analysts to collaboratively explore and analyze information within evidence chains;

- Focused research on hypothesis generation to identify chokepoint activities;

- Investigation of methods to assist in the exploitation of open source data; and

- Stabilization of the KDD prototype test bed for evaluating research components and their operational value.

Evidence Chaining:

Evidence chaining is a powerful method of achieving useful representation of information in its own right, but it has two additional advantages for the cross agency data sharing. First, it allows the collaborative information space to be constrained to only the information and entities in a neighborhood of the entities of interest. Second, it allows the data owners to place limits on the data that they will share. In KDD, rule based filters between the data owners and the shared information space are explored. The process of accumulating new information through policy and rule based filters and introducing them to find additional related information in numerous databases is commonly referred to as "Dipping".

Although there are systems engineering issues that are needed to make KDD's "Dipping" more efficient and more generic, this process appears to be an effective process to accumulate and organize data relevant to a particular intelligence target if seeded with one or more entities related to the target. This reduces the information space from billions of records in the distributed databases to thousands. Even with this reduction,

the resulting information space is often still too large and complex to be easily analyzed. A focus of KDD research in FY-2006 and beyond is to develop tools that will explore techniques to improve analysts' ability to collaboratively explore and analyze this space.

Hypothesis Generation:

One of the areas of research pursued in KDD has been to compare the evidence accumulated about a target against a formally or informally generated set of possible/hypothetical activities of interest. If one could reproducibly and reliably do this it would serve as fundamental components of an inference engine for large and complex information. The "Hypothesis Space" is essentially a work breakdown structure for a particular activity of interest to the level of observable evidence. Therefore, if the target were engaged in the activity, they would generate a particular pattern of evidence that is detectible. Conversely, if we observed a particular pattern of evidence then we have evidence consistent with an activity of intelligence interest.

Although the potential of this idea is substantial, ongoing research discovered two challenges that suggested a substantial change is required in the direction of this work. First, the converse of the proposition would likely generate many false alarms, particularly in applications where "dual use" activity was significant. Second, "Red Teaming exercises illuminated the often the generation of observables for an entire activity was unwieldy and nearly impossible to fully generate.

From this lesson learned, there has been a change in the direction of the research. Although an activity of interest may have a very large breakdown structure, there are specific sub-activities that are often well defined and suspicious. Substantial portions of the activities of interest include several chokepoint activities along with other activities that are more or less unconstrained. FY-06 research and beyond must focus hypothesis efforts on these chokepoint activities.

Open Source:

The importance of open source information to intelligence analysis and to the law enforcement community has been significant for some time. The Intelligence Community has placed increased emphasis on open source information as a critical intelligence source and has established a new management structure to address this particular discipline. Much of the KDD research has applicability to the analysis of open source information. The KDD Program will increase its emphasis on testing KDD tools in the prototype system using information from a wide variety of publicly available data sources. This emphasis on working with open source material from a diverse set of data sources should assist KDD funded researchers to test new approaches against massive data sets within the KDD prototype architecture of BLACKBOOK.

Stabilizing BLACKBOOK and Test and Evaluation:

By the end of 2005, the KDD Program Office will disseminate a more professionalized version of the KDD prototype software, BLACKBOOK, along with a well-structured tutorial on how to use the system and integrate research results within the framework. Consequently, during 2006 the KDD Program Manger expects that KDD funded researchers will employ the KDD prototype in their research and deliver new capabilities within the BLACKBOOK framework architecture. This will allow the products of KDD research to be evaluated in a timely manner with real intelligence problems without extensive modification. Upon signature of a no-cost license agreement by an authorized official of the research university, a copy of the KDD prototype system and documentation will be supplied to the researcher. The KDD Program Office is investigating the legality of issuing the BLACKBOOK software and supporting documentation through a special researchers Internet portal that can be accessed using a special password by those who have signed a no-cost license agreement. Until the portal is operational, BLACKBOOK will be distributed via mail on a CD.

In late 2005, the KDD Program Manager will open the new Multi-INT Program Center (MIPC) at Johns Hopkins Applied Physics Laboratory in Laurel Maryland. In the MIPC, the KDD team will test and evaluate data fusion and knowledge discovery technology; the results of KDD funded research; and stage deployments of tests of the KDD prototype system with operational customers. The MIPC will be collocated with a DHS laboratory and will be connected to a number of DoD and Intelligence Community agencies. Establishment of the MIPC is commensurate with an increase in the size of the programmer and systems engineering team devoted to the KDD Program. Consequently, the level of interaction between the KDD Program staff members and KDD researchers will intensify, as more research deliverables are available, tested, and evaluated.

KDD Process and Timeline For Funding Research Activities

The annual KDD Conference begins the annual research process cycle. At this Conference, researchers present the results and status of their research activities that have been funded by the KDD Program. The KDD Program Manager issues a call for and establishes a specific deadline date for submission of White Papers (brief -5 page submissions that outline the nature of the research proposed for the next year). The White Papers are submitted electronically in Microsoft Word to the KDD Program Office by the designated date. The deadline for submission of White Papers is approximately 3 weeks following the annual conference.

The KDD Working Group and the KDD Program Manager review the White Papers. This process will take about 30 days. Following the review of the White Papers, the KDD Program Manager will invite a select number of researchers to submit a Proposal for KDD funded research. The call for proposals usually occurs within 60 days of the KDD Conference. All will be notified of the specific dates at the annual conference and via email following the conference.

Proposals for KDD research shall be submitted electronically to the KDD Program Office (dlricha@afterlife.ncsc.mil) with a courtesy copy to ahbecke@prodigy.net. The format for the proposal is 6-10 pages text with 1 additional page that identifies the cost and level of manpower associated with each specific task contained in the proposal. A proposal may address a multi-year research effort however costs for each year, manpower level for each task, and specific deliverables must be separately identified. The proposal must describe the nature of the research proposed, the anticipated output of the research, each task to be accomplished, and an identification of specific deliverables. The KDD Program Manager will notify those researchers whose projects are selected for funding by the KDD Program. The KDD Program Manager will also notify those researchers whose projects will not be funded and indicate to the extent possible the reason the project is not being funded. Proposals may not be submitted directly by contractors or companies. Contractors and companies must submit research proposals through a government agency that is sponsoring the research. KDD funds will be distributed through appropriate agencies and departments for the various research activities. For example, those projects funded at universities will receive their KDD funding through the National Science Foundation (NSF). Government agencies will be funded directly through money transfers from ITIC. Companies that are supporting research on behalf of government agencies and departments will receive their KDD funding through the agency or department with which they have a contract.

Researcher Responsibilities

Researchers who receive KDD funding shall agree to:

- Develop all deliverables including software, algorithms, etc. in the KDD prototype system framework, BLACKBOOK.

- Deliver the results of their research to the Project Monitors in CD format. The KDD Project Monitors are responsible for delivering the research results to the KDD Program Manager.

- Meet at least twice annually with their Project Monitors

- Brief the results/status of their research at the annual KDD Conference

- Submit at least one technical article for publication in the Journal of Intelligence Community Research and Development (JICRD)

- Submit through the Project Monitors quarterly status reports on their KDD funded research projects and, as required by the KDD Program Manager, submit aperiodically, information on the status of their projects.