

ITR/IM+SII: A Distributed Information Management Framework (REGNET) for Environmental Laws and Regulations

Kincho H. Law¹

Professor of Civil and Environmental Engineering
Stanford University
Stanford, CA 94305-4020
Email: law@stanford.edu

ABSTRACT

The complexity, diversity, and volume of Federal and State regulations (as well as supplementary and supportive documents) are detrimental to businesses and hinder public understanding of government. The objective of REGNET project is to develop information infrastructure and tools for regulatory information management and to facilitate compliance assistance. As a pilot research application, the REGNET project focuses on environmental regulations. The basic research tasks include: (1) textual parsing and storage, (2) semi-structured, indexed storage, (3) means to resolve semantic ambiguities, (4) cross-referencing appropriate for automated retrieval and analysis of relevant documents, and (5) on-line compliance checking of governmental regulations. The experimental scope of this project focuses on Code of Federal Regulations (CFR) Title 40: Protection of the Environment and California Code of Regulations (CCR) Title 22: Social Security. Implementation examples include regulations and selected supplementary documents, covering hazardous waste, drinking water and the management of used oil.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *retrieval models*, I.2.1 [Artificial Intelligence]: Applications and Expert Systems – *law*, I.7.1 [Document and Text Processing]: Document and Text Editing – *Document management*.

1. INTRODUCTION

There has been a push by the executive office that government agencies put more emphasis on compliance assistance in lieu of enforcement to encourage companies to comply with regulations. It is well recognized that the complexity, diversity, and volume of Federal and State regulations are detrimental to businesses and also hinder public understanding of government. In addition to the regulations, supplementary and supportive documents (such as preambles, interpretation guides) are also an important part of regulatory information. The objective of REGNET project is to develop a formal information

infrastructure for regulatory information management and to facilitate compliance assistance. As a pilot research application, the REGNET project focuses on environmental regulations. The basic research tasks include: (1) textual parsing and storage, (2) semi-structured, indexed storage, (3) means to resolve semantic ambiguities, (4) cross-referencing appropriate for automated retrieval and analysis of relevant documents, and (5) on-line compliance checking of governmental regulations. The experimental scope of this project focuses on Code of Federal Regulations (CFR) Title 40: Protection of the Environment and California Code of Regulations (CCR) Title 22: Social Security. Implementation examples include regulations and selected supplementary documents, covering hazardous waste, drinking water and the management of used oil.

2. CURRENT STATUS

2.1 Repositories and Access Tools

A textual parser has been designed and implemented to parse online HTML-based federal and state (California) environmental regulations into an XML structure. Specifically, the parser was successfully applied to parse the entire CFR 40 on Protection of Environment and CCR 22 on Social Security (as well as CRF 21 on Food and Drugs and many other regulations.) In addition, the repository also includes selected supplementary documents dealing with used oil, which include the preamble to the regulation text found in 40 CFR 261 and 279, administrative decisions, guidance documents, federal cases, letters from the general counsel and letters of interpretation from the EPA. Online access tools have been built to allow searching and retrieval of regulations and documents.

2.2 Ontology Development and References

One key issue dealing with a voluminous set of regulatory documents is in to build appropriate ontology and concepts that can facilitate linking related regulations and documents stored in the repository. A set of tools have been developed to extract vocabulary and semi-automatically build structural thesauri (ontologies) for regulatory domains of interests. A parsing system has been developed using a context-free grammar and a semantic representation/interpretation system to automate the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference'04, Month 1–2, 2004, City, State, Country.
Copyright 2004 ACM 1-58113-000-0/00/0004...\$5.00.

¹ Grant No.: EIA-0085998

CO-PI: Gio Wiederhold, Jim Leckie and Barton Thompson
Research Assistants: Shawn Kerrigan, Gloria T. Lau, Bill Labiosa, Charles Heenan, Haoyi Wang, Xiaoshan Pan, Liang Zhou, Pooja Trivedi, Jun Peng
Institution: Stanford University
Website: <http://eil.stanford.edu/regnet>

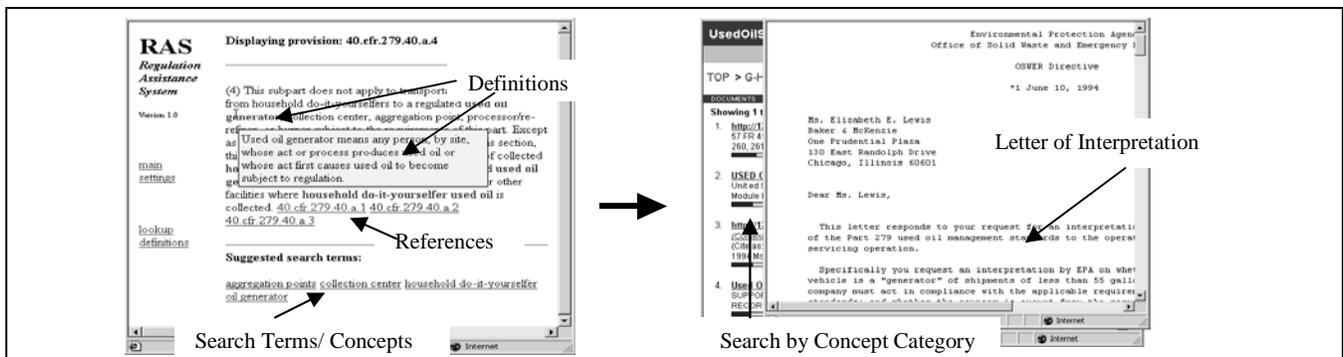


Figure 1: Definition, reference and concept tags



Figure 2: Linking specific application to the compliance assistance system

extraction of references and to tag regulation provisions with the list of references they contain. A research tool has also been developed to extract definitions on domain-specific terms and acronyms and attached them to the terms. Figure 1 shows the concept, reference and definition tags generated using the tools developed.

2.3 On-Line Compliance Checking

An important objective of this research is to provide the means to interface the regulations with usages such that the regulations are not passive but active documents that can be dynamically linked to application programs for users to search and access regulations and to perform compliance checking. Expressing the rules in terms of First Order Predicate logic sentences and adopting (publicly available) theorem provers, a regulation assistance system has been designed and implemented to assist the compliance of provisions related to used oil management (40CFR260 and 40CFR279). The web-based compliance assistance system helps guide the user through the regulations, automatically insert links to any referenced regulation provisions, display terms and definitions and enable instant access to repository documents related to the provision. One can build a compliance system for specific application utilizing the RCC system and the repository as a back end. Figure 2 illustrates a demonstration system built for vehicle maintenance shops to check compliance with the used oil provisions.

3. RESEARCH PLAN

As noted, legal regulations and information arise from diverse sources, each source has its own objectives, semantics, documentation format, and organization. We have begun to investigate a knowledge composition approach and develop a tool to extract features and to compare similarities to determine the “relatedness” of regulations and documents. As a

demonstration, similarity has been discovered between sections from Parts 141-143 of 40 CFR and Division 4 of 22 CCR on regulations related to drinking water. The prototype system has also been applied to a E-rulemaking scenario where the proposed provisions on “Guidelines for Accessible Public Right of Way,” issued by the Access Board were compared with over 1400 public comments to identify their relatedness. This preliminary experimental study has revealed the potential of a knowledge-driven similarity analysis approach to regulatory information management. The plan is to further develop and validate the “relatedness” analysis approach developed in this study and to demonstrate this fundamental approach for a broader set of federal and state environmental regulations, namely hazardous waste management. Furthermore, new application of the relatedness analysis framework to develop “regulatory locator” for specific domains will be investigated.

4. COLLABORATION AND OUTREACH

While this research focuses on the development of IT framework for regulatory information management and compliance assistance, the researchers have also been actively participating in the social and legal aspects of regulation compliance, enforcement and rule-making process. The researchers have participated in a number of workshops sponsored by Stanford’s Law School, Harvard’s Kennedy School of Government, EPA’s Office for Enforcement and Compliance Assurance related to IT and regulations. Active collaborations have been initiated with EPA, USGS, Pacific NW National Lab, Access Board and other agencies. These activities have led to the better understanding of the complexity in regulatory processes and the urgent needs of IT tools that could be useful by the public.

Project Title:

*ITR/IM+SII: A Distributed Information Management
Framework (REGNET) for Environmental Laws and
Regulations*

Grant No.: EIA-0085998

Principal Investigators:

**Kincho H. Law, Gio Wiederhold, James Leckie and Barton
Thompson**

**Department of Civil Engineering, Department of Computer
Science and Law School**

Stanford University

Stanford, CA 94305-4020

Collaboration with Government Agencies

- Environmental Protection Agency
 - Region 9 Office (Cheryl Nelson, Kevin Mayer)
Role: provided advice and feedback
 - Office of Enforcement and Compliance Assurance, OECA (Emily Chow and Jim Edwards)
Role: provided advice and feedback, provided linkage to research activities, sponsored technical session at agency's forum attended by professionals and agency's personnel, coordinated meetings with government laboratories
- USGS (James Rytuba, Richard Bernknopf)
Role: consultation and collaboration, provided student support during summer
- Regional Water Quality Control Board for San Francisco Bay (Tom Mumley)
Role: consultation and collaboration, provided assistance on data collection

Collaboration with Companies, Individuals and Organizations

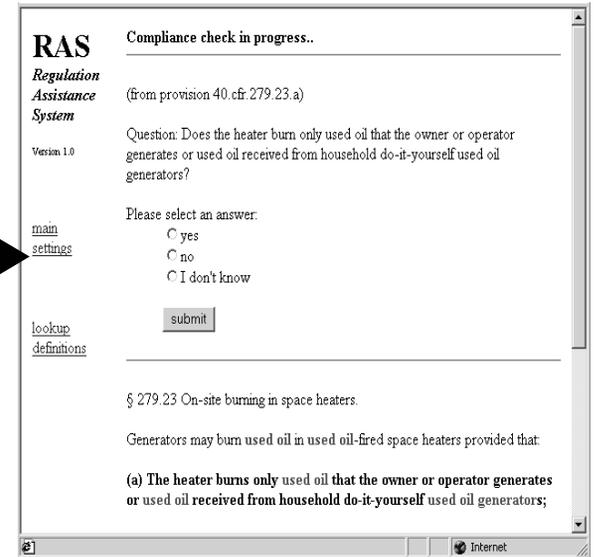
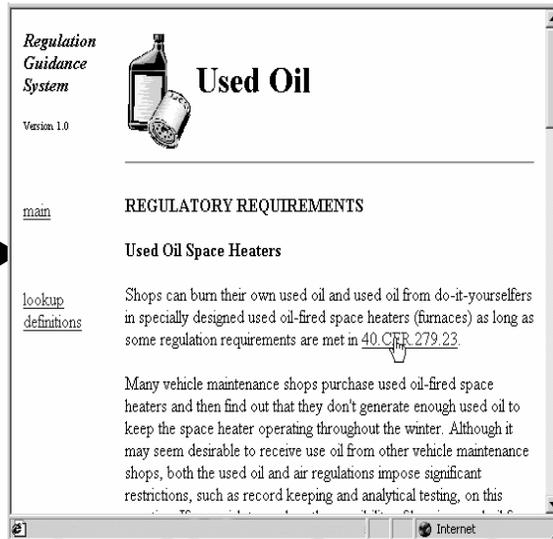
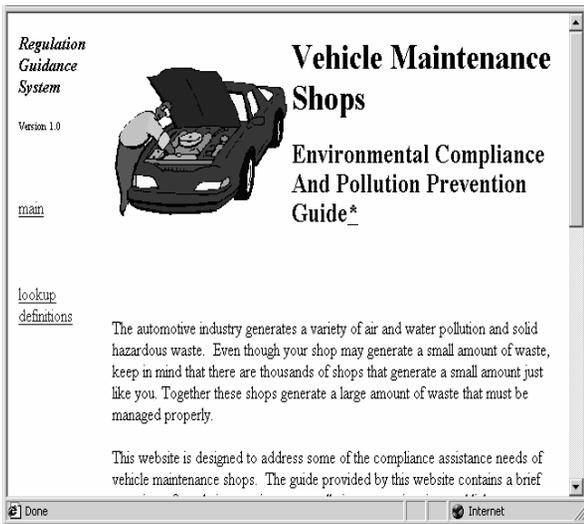
- Semio Corporation
Role: Software support (worth ~\$80,000) for research and development
- Intel Corporation
Role: Equipment grant (worth ~\$30,000 computers and servers) in support of research and development

Research Objective:

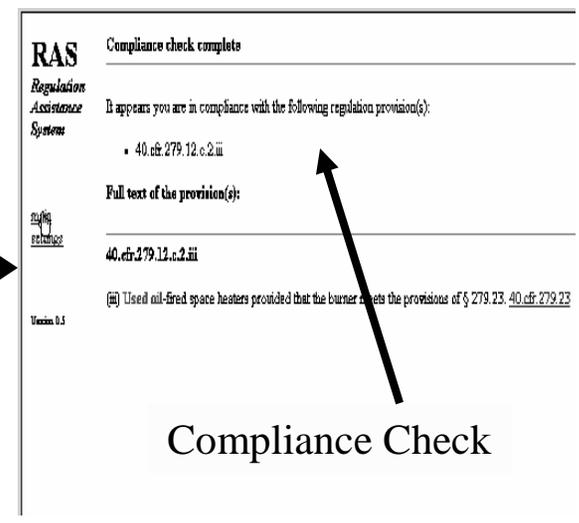
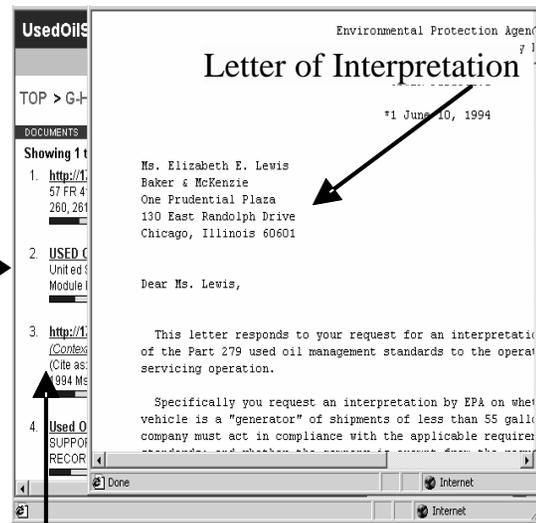
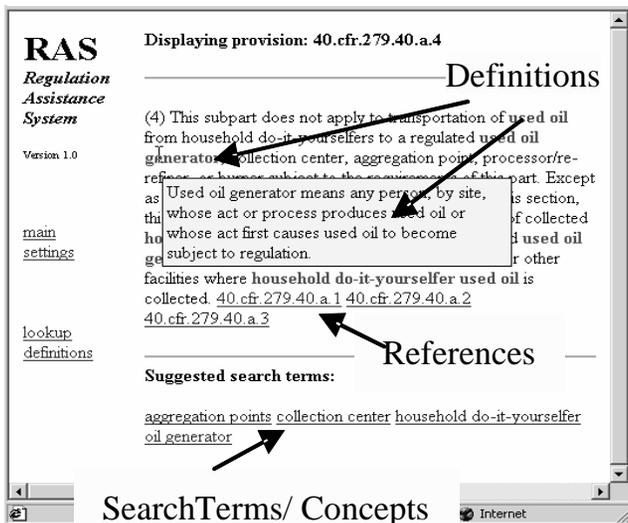
- **To develop information management frameworks that can facilitate public access to regulations, improve the efficiency of regulation compliance and facilitate the compliance process.**

Accomplishments:

- **Repositories:** Developed an XML-based framework for online repository of regulations and parsers to regulation texts into processable form
- **Access Tools:** Developed tools to access of the regulation text and related information (such as letters of interpretation, court cases, etc...)
- **Ontology Development:** Developed text mining tools to extract terms, concepts and their relationships in the regulations and among the different regulations and external documents
- **Integrated Access:** Developed information retrieval techniques to assist retrieval of regulations based on the content or relationships between the regulations
- **Analysis Tools:** Developed tools based on similar analysis techniques to compare regulations from different sources and support E-rulemaking process by linking public comments with proposed rules
- **Compliance Checking Assistance:** Developed a logic-based framework for compliance assistance and means to interface the regulations with usage.



Linking Specific Compliance Assistance Domain to RAS

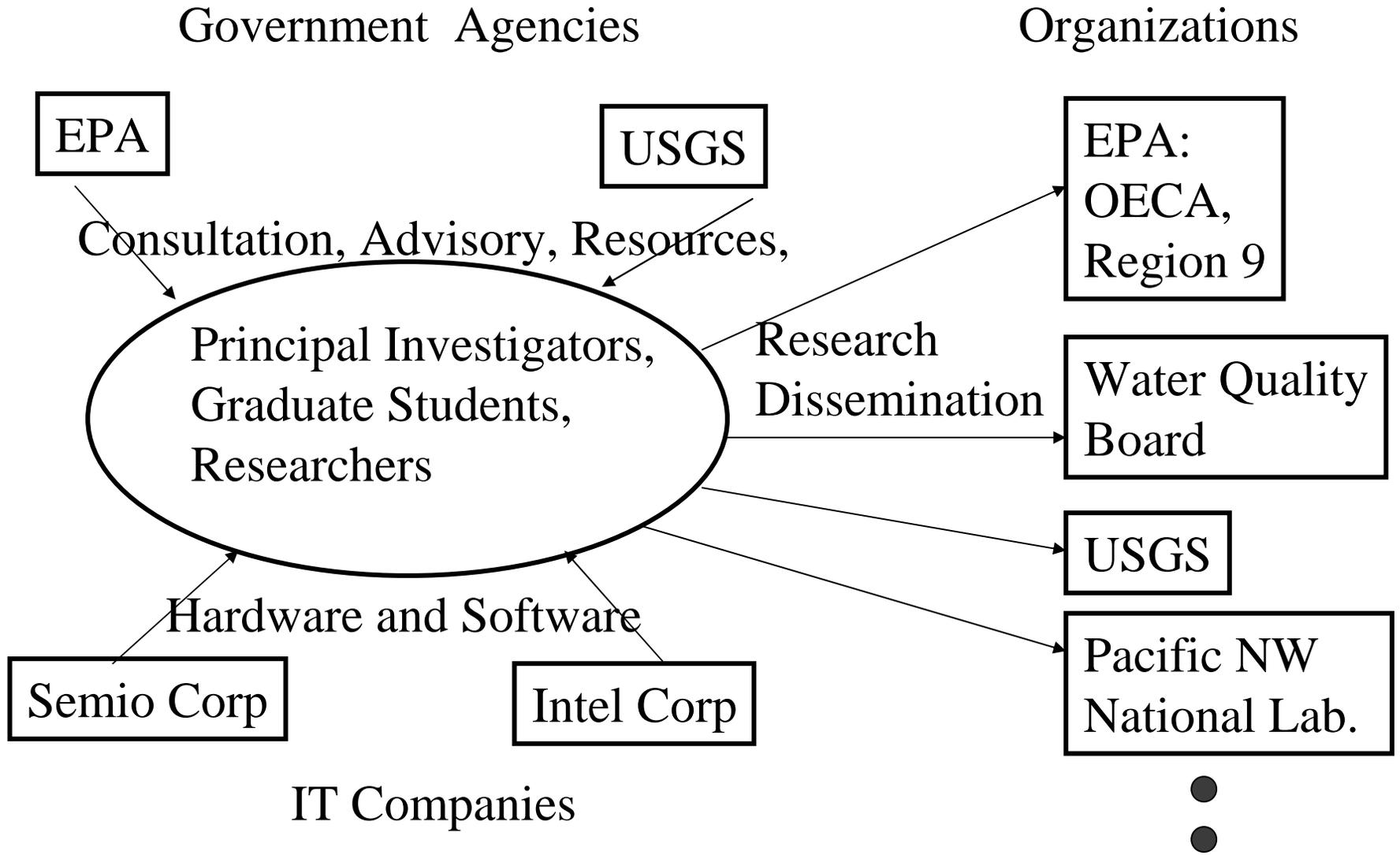


Search by Concept

Compliance Check

Regulations, Supplementary Document and Compliance Results

Management Structure



Collaboration Examples

- Worked collaboratively with EPA Office of Enforcement and Compliance Assurance -- sponsored technical session at EPA Compliance Forum
- Worked collaboratively with United State Geological Survey (USGS) - - sponsored graduate student for summer research
- Worked collaboratively with Water Quality Control Board -- provided assistance on data collection
- Dissemination of research results to EPA's OECA and Pacific NW National Laboratory
- Software support from Semio Corp. as well as equipment grant from Intel in support of research
- Worked closely with researchers at other universities

Broad Impact

- **Engineering Related Development:**
 - Develop tools for better understanding of environmental regulations and legal impacts
 - Develop tools to assist compliance of environmental regulations
- **Computer Science Related Development:**
 - Develop framework for the management of semi-structured documents, information access and retrieval of regulatory information as well as related legal documents
 - Develop text mining and similarity analysis techniques for semi-structured document analysis
 - Develop and demonstrate a logic-based framework for compliance assistance
- **Social and Legal Related Issues:**
 - Study legal implication of regulations and the role and impact of IT in regulation management
 - Study potential use of IT for supporting E-rulemaking and rule compliance processes
- **Education:**
 - Conduct interdisciplinary research between computer science, civil engineering and law
 - Students from computer science, engineering and law working together and gaining interdisciplinary research experience
 - Lectures in Freshman and graduate seminar courses related to IT, E-business and law

Barrier:

- Difficult to solicit supplemental (monetary) funding from government agencies on fundamental research beyond in-kind human resources and miscellaneous supports (such as summer internships and collaborations)

Opportunities:

- Transfer of new concepts to government agencies, laboratories and private companies to better serve the public
- Build bridges between academic research, government services and businesses
- Opportunity to develop non-traditional educational and research training involving engineering, computer science and law

Value of Digital Government Research

- **Research**

- To conduct research relevant to governmental needs and, in turn, benefit public access and enhance public understanding of government services
- To develop and demonstrate innovative research and developments using IT to support government related activities

- **Education**

- Provide opportunities to educate students about government services and regulations
- Encourage collaboration among different disciplines to conduct research and education related to government activities
- Develop new interdisciplinary research program to bridge the gap between theory and applications

Recommendations for Digital Government Program

- To continue supporting research projects that can bridge the state of the art IT academic research and practice in governmental activities
- To sponsor workshops and forum between academic researchers and government agencies regarding the future needs of the agencies
- To play an active role in technology transfer, from research to development to practice