

LITERATURE-RELATED DISCOVERY AND INNOVATION

DR. RONALD N. KOSTOFF

GEORGIA INSTITUTE OF TECHNOLOGY

RESEARCH AFFILIATE

SCHOOL OF PUBLIC POLICY

rkostoff@gmail.com; ronald.kostoff@pubpolicy.gatech.edu

PRESENTED: GEORGE MASON UNIVERSITY

BIOINFORMATICS AND COMPUTATIONAL BIOLOGY
COLLOQUIUM

20 MARCH 2012

OVERVIEW

- What is literature-related discovery and innovation (LRDI)?
- What are its main capabilities?
- Illustrative examples
 - Knowledge discovery and characterization
- What are its potential benefits?
- Who are its potential sponsors?
- What are future initiatives
- What are some of the studies/accomplishments?
- Backup

WHAT IS LRDI?

- LRDI is the extraction of useful information from text
- Digital text format
- Electronic extraction
- Large volumes of material

The techniques presented today will allow discovery and innovation to be generated systematically for any problem and any sponsor

MAIN LRDI CAPABILITIES

(operational and functional components)

- Two operational components
 - information retrieval
 - retrieval analysis
- Two main functional components
 - Discovery (*Link disjoint literatures for value-added*)
 - Scientific/technical/medical/knowledge discovery
 - literature-based discovery
 - literature-assisted discovery
 - Characterization (*Provide snapshot of technology*)
 - single technology core assessment
 - single technology core and expanded assessment
 - 4 • country assessment

DISCOVERY CAPABILITIES

- Links multiple literatures
- Can answer following types of discovery questions:
 - *Medical Discovery and Innovation*
 - What are potential treatments for Extensively Drug-Resistant Tuberculosis
 - How can we improve immunity to West Nile Virus
 - *Scientific Discovery and Innovation*
 - How can we improve water purification technology
 - How can we improve transport and distribution logistics
 - *Knowledge Discovery and Innovation*
 - How can we assess a country's weapons development capability
 - How can we improve logistics planning for emergency evacuations
 - **EXAMPLES AT END OF PRESENTATION**

DISCOVERY CAPABILITIES (CONT'D)

- **MISSION ASSURANCE:** What are **impacts of emerging technologies on cyber security, information assurance?**
- **HOMELAND SECURITY:** What are alternative network structures for infrastructure survivability?
- **BIOSECURITY:** What are impacts of emerging technologies on biosecurity and health?
- **CLIMATE CHANGE:** What are impacts of climate change on health and biosecurity?
- **VETERANS AFFAIRS:** What are potential treatments for brain injury, spinal cord injury?
- **EMERGING TECHNOLOGIES:** What are recent early-stage emerging technologies? What is the full spectrum of their potential impacts?
- **HEALTHCARE:** What are potential treatments for infectious and chronic diseases?
- **COMPREHENSIVE SCREENING:** Which emerging technologies could have substantial impact on improving comprehensive screening?

CHARACTERIZATION CAPABILITIES

- Quantifies research infrastructure and discipline efforts
- Can answer following types of questions:
 - **How do the quantity and quality of China's publications compare to that of the USA**
 - **What are the main sub-themes in global nanotechnology research**
 - What is the collaboration pattern among China's research institutions and with external research institutions
 - What is the global infrastructure (prolific researchers, institutions, journals, countries, etc) of the information assurance research literature
 - **SEE BACKUP SLIDES FOR USA/CHINA COMPARISON**

LRDI'S POTENTIAL BENEFITS

- Provides powerful capability to support multiple organizations in multiple tasks
 - identifies global S&T
 - allows strategic planning of site visits
 - allows leveraging/coordination
 - identifies emerging technologies
 - prevents technological surprise
 - identifies potential new research directions
 - identifies scientific discovery
 - identifies networks of linked entities
 - intelligence

POTENTIAL SPONSORS

- Homeland security
- Defense
- Health
- Energy
- EPA
- All Federal agencies

POTENTIAL INITIATIVES

- **GRAND CHALLENGES**

1. Energy – Insure an uninterrupted supply of adequate low-cost clean energy.
2. Water – Insure an adequate supply of water for drinking, agriculture, and other critical uses.
3. Population – Control world population to a sustainable level.
4. Space – Exploit the potential of near-earth and outer space for human benefit.
5. Environment – Minimize environmental damage, especially due to climate change.
6. Transportation – Transport people and materiel inexpensively, rapidly, and safely to allow a reasonable standard of living while minimizing damage to the environment.
7. Health – Improve health and extend longevity for high quality long life.

POTENTIAL INITIATIVES (CONT'D)

- **GRAND CHALLENGES (CONT'D)**

8. Education – Improve education to maximize each person's potential while achieving larger social goals.
9. Brain – Exploit the processes of the brain to solve societal problems.
10. Food – Insure an adequate, safe, low cost, and clean food supply.
11. Security – Enhance national and personal security.
12. Infrastructure - Protect and improve the national infrastructure.
13. Waste – Insure that all forms of waste are kept at manageable levels cheaply and safely.
14. Discovery – Accelerate discovery and innovation.
15. Information – Maximize use of information technology to reduce physical resource requirements and to address the above fourteen challenges.

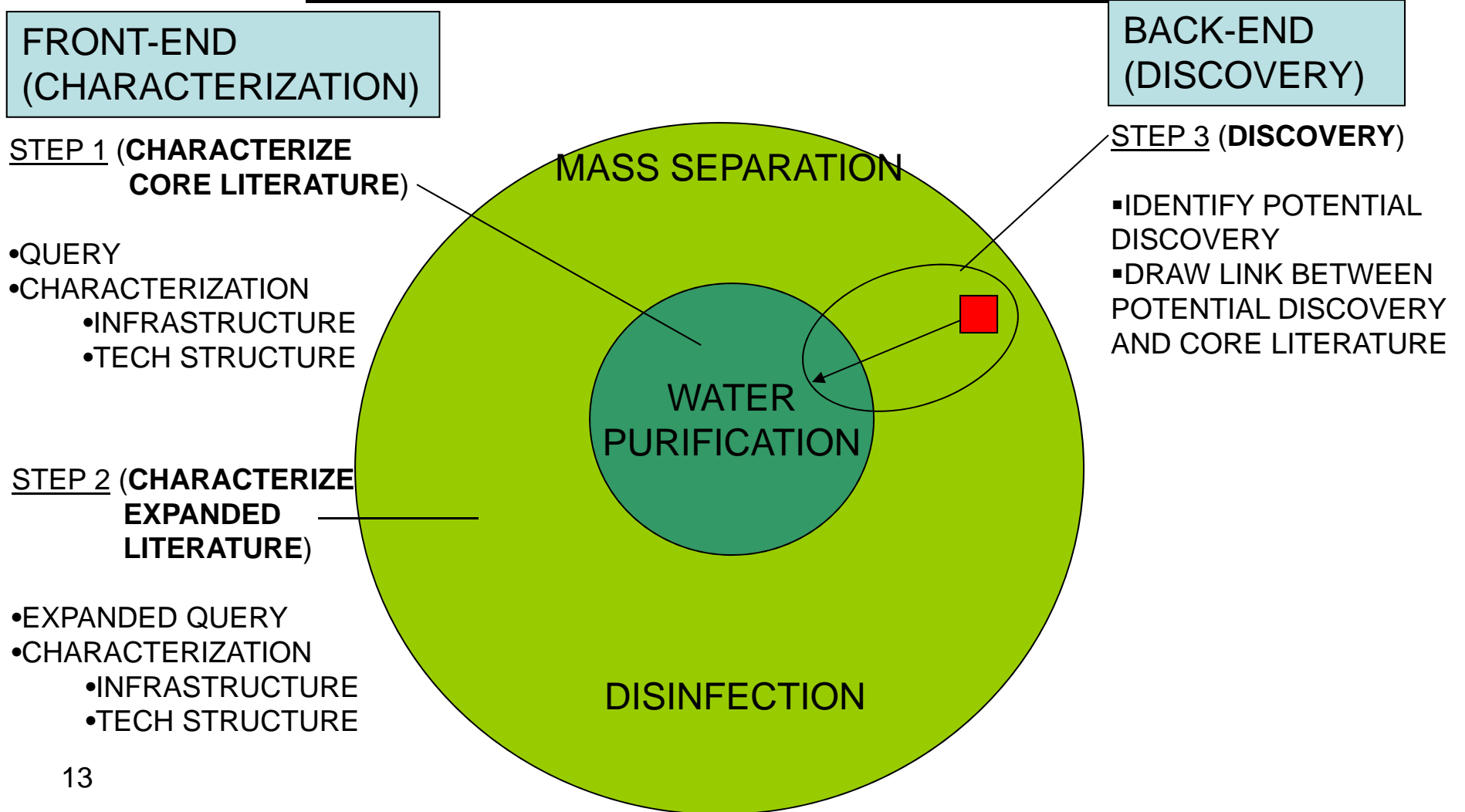
LRDI CAPABILITIES (*DISCOVERY*)

- **Discovery background**

- Discovery and innovation critical for modern economies and militaries
- Radical discovery requires insights from disparate disciplines
- Increased specialization reduces awareness of other disciplines
- *Require method for systematic access to other disciplines*

RADICAL DISCOVERY AND INNOVATION

(INSIGHTS FROM DISPARATE LITERATURES)
SEE NEXT SLIDE FOR NARRATIVE



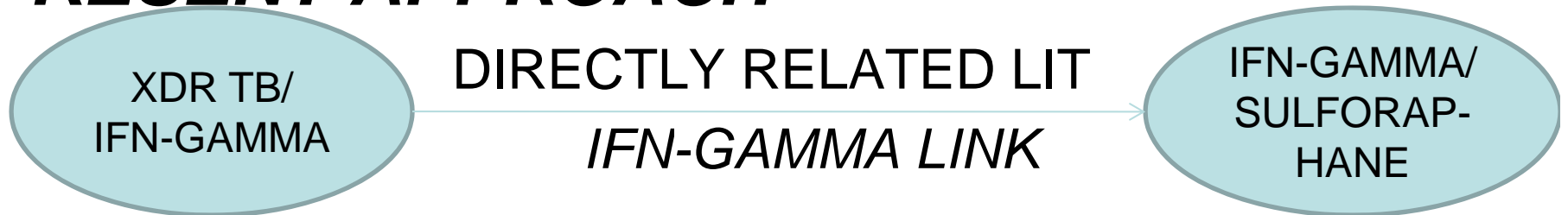
RADICAL DISCOVERY AND INNOVATION *NARRATIVE*

- Retrieve core literature for problem of interest (e.g., water purification)
- Identify thrust areas by clustering/factor analysis
- Retrieve literatures related to thrust areas that present technical/medical challenges (e.g., mass separation)
- Subtract core literature from expanded literature
- Identify potential discovery candidates
- Relate potential discovery to core literature challenges

DISCOVERY APPROACH – SCHEMATIC

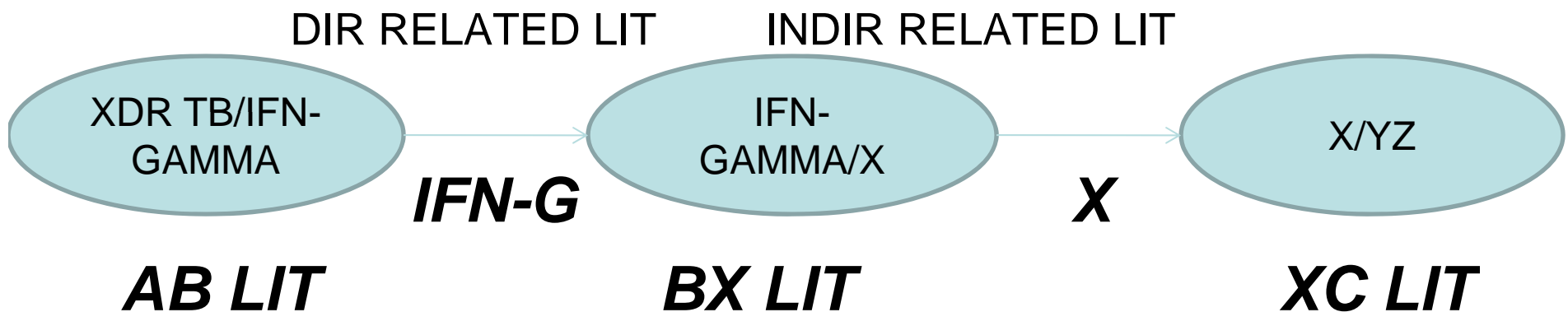
NEXT SLIDE FOR NARRATIVE

- **PRESENT APPROACH**



- AB LITERATURE
- BC LITERATURE
- B PHRASES LINK DISJOINT LITERATURES A,C

- **PROPOSED APPROACH**



DISCOVERY APPROACH – CONT'D

- DISCOVERY FROM DIRECTLY-RELATED LITERATURES
 - EX. INCREASED INTERFERON GAMMA ASSOCIATED WITH REDUCED EXTREMELY DRUG RESISTANT TUBERCULOSIS
 - INCREASED INGESTION OF SULFORAPHANE INCREASES IFN
 - THEREFORE, INCREASED SULFORAPHANE TREATS XDR TB BY INCREASING IFN

- DISCOVERY FROM INDIRECTLY-RELATED LITERATURES
 - EX. INCREASED INTERFERON GAMMA ASSOCIATED WITH REDUCED EXTREMELY DRUG RESISTANT TUBERCULOSIS
 - INCREASED INGESTION OF SUBSTANCE X INCREASES IFN
 - INCREASED INGESTION OF SUBSTANCE YZ INCREASES SUBSTANCE X
 - THEREFORE, INCREASED INGESTION OF YZ TREATS XDR TB BY INCREASING X, WHICH IN TURN INCREASES IFN

DISCOVERY APPLICATIONS

- Literature-based discovery
 - Analyst performs front-end/ back-end
- Notifications (BAA, SBIR, etc)
 - BAA notification sent to experts identified in front-end
- Workshops
 - Experts identified in front-end invited to workshops
- Roadmaps
 - Experts identified in front-end form roadmap development teams

DISCOVERY APPLICATIONS (CONT'D)

- Notifications (journals)
 - special issue notification sent to experts identified in front end
- Advisory panels
 - experts identified in front end invited to participate in advisory panels
- Review panels
 - experts identified in front end invited to participate in review panels
- Points of contact
 - experts identified in front end serve as points of contact
- Organization and team structuring
 - experts and disciplines identified in front end used to structure teams and organizations

RECENT EXAMPLES

- MILITARY RELEVANT TECHNOLOGIES-2010
- PARKINSON'S DISEASE - 2008
- MULTIPLE SCLEROSIS - 2008
- SARS - 2009
- PARKINSON'S-CROHN'S - 2009
- VITREOUS RESTORATION - 2012
- LRDI SUMMARY PAPER - 2012

MILITARY RELEVANT TECHNOLOGIES

- IDENTIFY TECHNOLOGIES/RESEARCH THRUSTS THAT ARE MILITARY RELEVANT
 - 1A. IDENTIFY MILITARY IDENTIFIABLE TERMS (E.G., ARTILLERY, FIGHTER AIRCRAFT, ETC)
 - 1B. USE RELEVANCE FEEDBACK APPROACH TO GENERATE QUERY OF SUCH TERMS
 - 2A. SEARCH FOR TECHNOLOGIES IN RETRIEVED RECORDS CLOSELY ASSOCIATED WITH MILITARY APPLICATION
 - 2B. REPEAT PROCESS TO IDENTIFY RESEARCH THRUSTS CLOSELY ASSOCIATED WITH MILITARY RELEVANT TECHNOLOGIES
 - STUDY ONLY IDENTIFIED TECHNOLOGIES (2A)

MILITARY RELEVANT TECHNOLOGIES

- POINT TO EMPHASIZE
 - CANNOT TYPICALLY IDENTIFY THESE TECHNOLOGIES USING DOCUMENT CLUSTERING OR FACTOR ANALYSIS
 - THESE METHODS TEND TO PROVIDE **DISCIPLINE** BREAKDOWNS
 - MILITARY-RELEVANT, SPACE-RELEVANT, INTELLIGENCE-RELEVANT HAVE **APPLICATION** ORIENTATION

MILITARY RELEVANT TECHNOLOGIES

- STEPS 1A AND 1B WERE RELATIVELY STRAIGHT-FORWARD; LARGE QUERY RESULTED
- STEPS 2A AND 2B PROVED TO BE FAR MORE DIFFICULT

MILITARY RELEVANT TECHNOLOGIES

- OBJECTIVE: IDENTIFY TECHNOLOGY TEXT PATTERNS THAT OCCURRED IN THE MILITARY-IDENTIFIABLE RECORDS
- ADD THESE TEXT PATTERNS TO A QUERY THAT WOULD RETRIEVE RECORDS CONTAINING MILITARY-RELEVANT TECHNOLOGIES.
- PROBLEM: MANY TYPES OF TECHNOLOGY TEXT PATTERNS COULD BE EXTRACTED FROM THE MILITARY-IDENTIFIABLE RECORDS
- ***THESE DIFFERENT TEXT PATTERNS RESULTED IN DIFFERENT STRENGTHS OF RELATIONSHIPS BETWEEN THE TECHNOLOGIES IN THE RECORDS RETRIEVED AND THEIR LINKAGES TO THE MILITARY APPLICATION.***

MILITARY RELEVANT TECHNOLOGIES

- 1. GENERAL TECHNOLOGY PHRASES (E.G., SIGNAL PROCESSING) – MILITARY NON-SPECIFIC
- 2. MORE DETAILED TECHNOLOGY PHRASES (E.G., RADAR SIGNAL PROCESSING) – STILL MILITARY NON-SPECIFIC
- 3. TECHNOLOGY COMBINATIONS (E.G., ‘SIGNAL PROCESSING’ AND ‘NEURAL NETWORKS’) – SOMEWHAT BETTER, BUT STILL NON-SPECIFIC
- 4. **TECHNOLOGY-FUNCTION COMBINATION (E.G., ‘SIGNAL PROCESSING’ AND ‘TARGET DETECTION’, ‘GENETIC ALGORITHMS’ AND ‘FEATURE EXTRACTION’) – HIGHLY MILITARY SPECIFIC**

PARKINSON'S DISEASE

- APPROACH: IDENTIFY KEY MEDICAL CHALLENGES FROM PD CORE LITERATURE
 - RETRIEVE ALL ARTICLES THAT CONTAIN PD; CLUSTER RETRIEVAL
- IDENTIFY OTHER LITERATURES THAT ADDRESS THESE CHALLENGES
 - DEVELOP QUERY TO RETRIEVE LITERATURES
- SUBTRACT PD CORE LITERATURE
- VALIDATE FOR POTENTIAL DISCOVERY
 - CHECK PRIOR ART AGAINST SCI/MEDLINE/PATENTS

PARKINSON'S DISEASE

- QUERY FOR DISPARATE LITERATURES
- 1. DIRECTLY RELATED
 - KEY OBSERVATION: RECORDS MOST RELEVANT TO POTENTIAL DISCOVERY TENDED TO HAVE MORE THAN ONE OF THE KEY QUERY MECHANISM TERMS (IDENTIFIED FROM THE CLUSTERS) IN THE MESH FIELD.
 - IDENTIFY KEY PHRASES FROM EACH CORE LITERATURE THRUST (~15 TOTAL)
 - GENERATE ALL COMBINATIONS OF THREE PHRASES FOR GENERAL PHRASES
 - GENERATE ALL COMBINATIONS OF TWO PHRASES FOR SPECIFIC PHRASES
 - INTERSECT WITH NON-DRUG TERMS
 - INSERT IN SEARCH ENGINE; RETRIEVE RECORDS

PARKINSON'S DISEASE

- 2. INDIRECTLY RELATED
 - SAME OBSERVATION AS BEFORE ABOUT MULTIPLE KEY WORDS IN ABSTRACT BEING MORE ASSOCIATED WITH POTENTIAL DISCOVERY
 - CLUSTER DIRECTLY RELATED RETRIEVAL
 - IDENTIFY KEY WORDS/THEMES FROM EACH CLUSTER
 - USE COMBINATIONS OF KEY WORDS TO REDUCE RETRIEVAL VOLUME AND SHARPEN RETRIEVAL
 - INTERSECT WITH NON-DRUG TERMS
 - INSERT INTO SEARCH ENGINE; RETRIEVE RECORDS

PARKINSON'S DISEASE

- Conclusions excerpt: synergy of lifestyle/ dietary practices that could be interpreted as anti-Parkinson. Along with non-discovery items such as less dairy, green tea, caloric restriction, blueberries, broccoli/broccoli sprouts, and lower temperature cooking are potential discovery items such as malanga extracts, kolaviron, isohumulones, brown algae, and Rhododendrum flavonoids.
- major disconnect between [mainstream] therapies and the therapies suggested by what has already been demonstrated in the core PD literature, much less what we have generated from the related literatures. The major medical Web sites (and journal reviews) present about a half-dozen drug treatment options for PD, and perhaps 3-4 surgical/invasive procedures

MULTIPLE SCLEROSIS

- SIMILAR APPROACH AS PD STUDY WITH ONE DIFFERENCE
- PD USED COMBINATIONS OF KEY PHRASES TAKEN FROM ALL CLUSTERS (INTER-CLUSTER)
- **MS USED COMBINATIONS OF KEY PHRASES TAKEN FROM EACH CLUSTER (INTRA-CLUSTER)**
- MOST INTER-CLUSTER COMBINATIONS PROVED OVERLY RESTRICTIVE
- INTRA-CLUSTER COMBINATIONS MADE MORE EFFICIENT USE OF LIMITED PHRASES

MULTIPLE SCLEROSIS

- **CONCLUSIONS:** picture from discoveries is a synergy of lifestyle/dietary practices that could be interpreted as anti-MS. along with non-discovery items such as vitamin D, dietary chelators, caloric restriction, complement-inhibitory herbs, nigella sativa oil, green tea, and quercetin are potential discovery items such as shogaol, ethanol, iron, petaslinolide a, mangifera indica l, tiliroside, gnaphaliin, cissus quadrangularis extract, kalpaamruthaa, salvia miltiorrhiza bunge, inchinko tj-135, silymarin, edaravone, sopoongsan, and artemesia iwayomogi.

SARS

- BACKGROUND LITERATURE SURVEY OF SARS LITERATURE SHOWED
 - 8000 PEOPLE PRESENTED WITH SARS
 - 10% SUCCUMBED
 - DRUG TREATMENTS DIDN'T WORK; GOOD HYGIENE/QUARANTINE STEMMED THE SPREAD
 - THOSE WHO DIED HAD WEAK IMMUNE SYSTEM; CO-MORBIDITIES
 - **STRENGTHENING IMMUNE SYSTEM BECAME DISCOVERY TARGET**

SARS

- DIFFERENT QUERY FORM USED
 - EXPRESSED DESIRED OUTCOME
 - E.G., ENHANCE HUMORAL IMMUNITY, RESTRICT VIRAL ENTRY
 - COMBINE WITH PROXIMITY SEARCH CAPABILITY
 - E.G., ENHANCE W/N “HUMORAL IMMUNITY”
 - ALLOWS FULL-TEXT SEARCHING; ‘SURGICAL’ RETRIEVAL

SARS

- FULL TEXT SEARCHING GREATLY INCREASED RETRIEVAL OF RELEVANT RECORDS
- SEARCHING THE FULL TEXT FIELD RELATIVE TO THE COMMONLY USED ABSTRACTS FIELD INCREASES RETRIEVALS BY ONE OR MORE ORDERS OF MAGNITUDE
 - FOR PHENOMENA-TYPE CATEGORIES (E.G., BLOOD FLOW, THERMODYNAMIC EQUILIBRIUM, ETC), RETRIEVALS ARE ENHANCED BY ABOUT AN ORDER OF MAGNITUDE.
 - FOR INFRASTRUCTURE-TYPE CATEGORIES (E.G., EQUIPMENT TYPES, SPONSORS, SUPPLIERS, DATABASES, ETC), RETRIEVALS ARE ENHANCED BY WELL OVER AN ORDER OF MAGNITUDE, AND SOMETIMES MULTIPLE ORDERS OF MAGNITUDE.
 - **FOR COMPETITIVE OR NATIONAL SECURITY INTELLIGENCE, MEMBERS OF THESE LATTER CATEGORIES AND THEIR RELATIONSHIPS TO THE MAIN THEMES BECOME OF HIGH IMPORTANCE**

SARS

- **CONCLUSIONS:** synergy of anti-SARS lifestyle/dietary practices. Along with non-discovery items such as urintricarboxylic acid (ATA), Emodin (an anthraquinone compound derived from genus *Rheum* and *Polygonum*), *Cimicifuga rhizoma*, *Meliae cortex*, *Coptidis rhizoma*, *Phellodendron cortex*, *Sophora subprostrata radix*, Betulinic acid, savinin, abietane-type diterpenoids and lignoids, quercetin-3-beta-galacto side, d.alpha,beta-unsaturated peptidomimetics, anilides, metal-conjugated compounds, boronic acids, quinolinecarboxylate derivatives, thiophenecarboxylates, phthalhydrazide-substituted ketoglutamine analogs, isatin, tannic acid, and 3-isotheaflavin-3-gallate (TF2B) are potential discovery items such as *Ganoderma lucidum*, Jacalin, Sulforaphane, methanolic extract of *M. koenigii* leaves, *Tinospora cordifolia*, Fucoidan, *Atractylodes macrocephala* Koidz, *Wasabia japonica*, seeds of *Sorghum bicolor* L, pidotimod and red ginseng acidic polysaccharide (RGAP), *Myrica rubra* leaf ethanol extract, *L. paracasei* NCC2461, methanol extract of *Asarum sieboldii*, Caffeoyl Glycoside, and adjuvants Korean mistletoe lectin C, *Cochinchina momordica* seed, Ag85B, probiotic *Bacillus cereus* var. *toyoi*, Kefir, Lentinan, polyphenol rich extract (CYSTUS052) from the Mediterranean plant *Cistus incanus*.

PARKINSON'S-CROHN'S

- PREVIOUS LRDI STUDIES WERE OPEN DISCOVERY SYSTEM (ODS)
 - START WITH PROBLEM; LOOK FOR SOLUTION
- PARKINSON'S-CROHN'S (PD/CD) WAS FIRST LRDI CLOSED DISCOVERY SYSTEM PROBLEM
 - START WITH TWO PROBLEMS; LOOK FOR LINKING MECHANISMS
- PD IS NEURODEGENERATIVE DISEASE; CD IS AUTOIMMUNE DISEASE
- IDENTIFY LINKAGES/COMMON FEATURES BETWEEN PD AND CD

PARKINSON'S-CROHN'S

- STANDARD CDS APPROACH (E.G., ARROWSMITH) LOOKS FOR COMMON PHRASES IN THE TWO DISPARATE PROBLEM LITERATURES
- TYPICALLY, MANY PHRASES IDENTIFIED; EACH PHRASE NEEDS TO BE VALIDATED
- RECORDS IN EACH LITERATURE THAT CONTAIN EACH PHRASE MUST BE READ FOR VALIDATION
- ARROWSMITH CONTAINS STATISTICAL FILTERS TO PRIORITIZE PHRASE SELECTION
- IS THERE A BETTER WAY TO PRIOTITIZE PHRASE SELECTION FOR LINKING LITERATURES?

PARKINSON'S-CROHN'S

- **COMBINED TWO APPROACHES FOR IDENTIFYING COMMON FEATURES IN TWO LITERATURES: TEXT-BASED AND CITATION-BASED**
- TEXT-BASED APPROACH IDENTIFIED RECORDS IN EACH LITERATURE THAT CONTAINED COMMON PHRASES (TITLE)
- CITATION-BASED APPROACH IDENTIFIED RECORDS THAT HAD COMMON REFERENCES
 - USED **BIBLIOGRAPHIC COUPLING**

PARKINSON'S-CROHN'S

- MATCHING SHARED PHRASES AND REFERENCES PROVIDED STRONG SYNERGY FOR PRIORITIZATION
- THREE MAJOR THEMES THAT UNIFIED THE PD AND CD LITERATURES
 - GENETICS; NEUROIMMUNOLOGY; CELL DEATH
- SOME NEW CONCEPTS AT THE SUB-SET LEVEL OF THE MAIN THEMES WERE IDENTIFIED
- EXAMINED TITLES AND ABSTRACTS
- FOR PROOF-OF-PRINCIPAL, USED ONLY TITLE PHRASES FOR PHRASE MATCHING; FAR MORE PHRASES AND INFORMATION CONTAINED IN ABSTRACT

VITREOUS RESTORATION

- COMPLETING A STUDY ON VITREOUS RESTORATION (VITREOUS IS THE GEL BETWEEN THE LENS AND THE RETINA IN THE EYE).
- USE IMPROVED VERSION OF THE FUNCTIONAL QUERY FIRST SHOWN IN THE SARS STUDY, INCLUDING PROXIMITY SEARCHING CAPABILITY.
- **USE A TEXT-BASED QUERY APPROACH IN CONCERT WITH A CITATION-BASED QUERY, WHICH EXPLOITS THE STRENGTHS OF EACH APPROACH AND ELIMINATES THE WEAKNESSES.**

VITREOUS RESTORATION

- MUCH TEXT MINING-BASED FILTERING FOR DISCOVERY RESULTED FROM THE SHARPNESS AND PRECISION OF THE QUERIES USED.
- REMAINING FILTERING WAS DONE THROUGH VISUAL EXAMINATION AND INSPECTION.
 - THE LATTER FILTERING STEP REQUIRED JUDGMENTS OF QUALITY, AND THAT EXCEEDED THE CAPABILITIES OF THE TEXT-BASED FILTERING APPROACHES.

VITREOUS RESTORATION

- REMOVED PREVIOUS RESTRICTIONS ON NON-DRUG NON-ADVANCED TECHNOLOGY DISCOVERIES ONLY; CONSIDERING ALL POTENTIAL FORMS OF TREATMENT.
- FINDING THAT GENERAL SYSTEMIC AND LOCAL PROBLEM-FOCUSED TREATMENTS ARE BOTH REQUIRED FOR OPTIMAL HEALING
 - TREATMENT EFFECTIVENESS WILL BE STRONGLY RELATED TO THE ABILITY TO IDENTIFY AND REMOVE CAUSES OF DISEASE.
- PLACING MORE EFFORT ON IDENTIFYING THE WIDEST SPECTRUM OF POTENTIAL CAUSES FOR VITREOUS DEGRADATION; INSUREs THAT POTENTIAL TREATMENTS COVER THE WIDEST SPECTRUM OF CAUSES POSSIBLE.
 - FINDING THAT A NUMBER OF POTENTIAL CAUSES HAVE NOT BEEN RESEARCHED IN THE LITERATURE, AND HAVE IDENTIFIED THESE AS RESEARCH GAPS.

VITREOUS RESTORATION

- PROBLEMS
 - IDENTIFY ALL POTENTIAL CAUSES OF VITREOUS DEGRADATION
 - NECESSARY TO ELIMINATE CAUSE BEFORE STARTING TREATMENT
 - SYNERGY OF MULTIPLE CAUSES
 - OPTIMAL COMBINATION OF TREATMENTS
 - HOW TO CULL LARGE NUMBER OF POTENTIAL TREATMENTS TO REALISTIC LIST
 - CREDIBILITY OF BIOMEDICAL LITERATURE
 - POOR RESEARCH; BIASED RESEARCH
 - INTERPRETATION OF CLINICAL TRIALS WITH MISSING VARIABLES

LRDI SUMMARY PAPER

- **COMPREHENSIVE AND PRECISE
INFORMATION RETRIEVAL IMPORTANT IN
DISCOVERY AND INNOVATION**
- **INTERDISCIPLINARY RESEARCH
EXTREMELY IMPORTANT IN DISCOVERY
AND INNOVATION**

LRDI SUMMARY PAPER

- BROAD BIOMEDICAL LITERATURE SEVERELY UNDER-UTILIZED FOR REVERSING CHRONIC DISEASE
- **MAJOR CONCERNS ABOUT THE CREDIBILITY AND INTEGRITY OF THE MEDICAL LITERATURE IN AREAS THAT CONCERN COMMERCIAL AND GOVERNMENT/POLITICAL SENSITIVITIES**

LRDI SUMMARY PAPER

- HORMESIS AND SYNERGY PLAY A CRITICAL ROLE IN PREVENTATIVE MEASURES AND ACCELERATED HEALING
- **CRITICAL NEED FOR CAUSE REMOVAL IN REVERSAL OF CHRONIC DISEASE**
- SEVERE UNDER-REPORTING OF CRITICAL VARIABLES IN THE CLINICAL TRIALS LITERATURE
 - CAUSE REMOVAL
 - PRACTITIONER SKILL
 - IMMUNE SYSTEM HEALTH
 - CIRCULATORY SYSTEM HEALTH

ILLUSTRATIVE MEDICAL DISCOVERY-1

(*Sulforaphane* for XDR TB Therapy)

- **Paper 1**: MDR-TB is often accompanied with the immunosuppression of the host. Given that we are unable to develop another potent anti-TB drug in near future, immunotherapy directed at combating immunosuppression and enhancing the host's own immune response is an attractive approach to supplement conventional chemotherapy for MDR-TB..Diverse cytokines are known to play an important role in anti-TB cell-mediated immunity, including IL-2, IL-12, IL-18 and **IFN-gamma**. Various animal experiments are indicating that administration of these cytokine(s) did recover the suppressed immunity and rescued the host from death by tuberculosis infection.Clinical trial of inhalation therapy with **IFN-gamma** showed some improvement for drug-resistant TB. **Cytokine treatment, however, often gave some deleterious side effects**.....
- **Paper 2**: Administration of **sulforaphane** significantly **enhanced the production** of Interleukin-2 and **Interferon-gamma** in normal as well as tumor-bearing animals.
- These data clearly suggest that **sulforaphane** effectively inhibited the spread of metastatic tumor cells through the stimulation of CMI, **upregulation of IL-2 and IFN-gamma**, and **downregulation of proinflammatory cytokines** IL-1beta, IL-6, TNF-alpha, and GM-CSF.

ILLUSTRATIVE MEDICAL DISCOVERY-2

(*Sulforaphane* for West Nile Virus Therapy)

- **Paper 1:** West Nile virus (WNV) causes a severe central nervous system (CNS) infection in humans, primarily in the elderly and immunocompromised. Prior studies have established an essential protective role of several innate immune response elements, including alpha/beta interferon (IFN-alpha/beta)..... we demonstrate that a lack of IFN-gamma production or signaling results in increased vulnerability to lethal WNV infection by a subcutaneous route in mice.....our experiments suggest that the dominant protective role of IFN-gamma against WNV is antiviral in nature, occurs in peripheral lymphoid tissues, and prevents viral dissemination to the CNS.
- **Paper 2:** Administration of *sulforaphane* significantly **enhanced the production** of Interleukin-2 and **Interferon-gamma** in normal as well as tumor-bearing animals.
- These data clearly suggest that *sulforaphane* effectively inhibited the spread of metastatic tumor cells through the stimulation of CMI, **upregulation of IL-2 and IFN-gamma**, and **downregulation of proinflammatory cytokines** IL-1beta, IL-6, TNF-alpha, and GM-CSF.

ILLUSTRATIVE SCIENTIFIC DISCOVERY-1

WATER PURIFICATION

(LESSONS FROM PLANT WATER FILTRATION)

- "It is shown how the complex, 'composite anatomical structure' of roots results in a 'composite transport' of both water and solutes. Parallel apoplastic, symplastic and transcellular pathways play an important role during the passage of water across the different tissues. These are arranged in series within the root cylinder (epidermis, exodermis, central cortex, endodermis, pericycle stelar parenchyma, and tracheary elements)... there is a rapid exchange of water between parallel radial pathways because, in contrast to solutes such as nutrient ions, water permeates cell membranes readily. The roles of apoplastic barriers (Casparian bands and suberin lamellae) in the root's endo-and exodermis are discussed. The model allows for special characteristics of roots such as a high hydraulic conductivity (water permeability) in the presence of a low permeability of nutrient ions once taken up into the stele by active processes... "

ILLUSTRATIVE SCIENTIFIC DISCOVERY-2

TRANSPORT AND DISTRIBUTION LOGISTICS

(VEHICLE ROUTING; PLANT NUTRIENT TRANSPORT LINKAGES)

- **Core paper:** The military "theater distribution vehicle routing and scheduling problem" (TDVRSP) is associated with determining superior allocations of required flows of personnel and materiel within a defined geographic area of operation. A theater distribution system is comprised of facilities, installations, methods, and procedures designed to receive, store, maintain, distribute, and control the flow of materiel between exogenous inflows to that system and distribution to end-user activities and units within the theater. An automated system that can integrate multimodal transportation assets to improve logistics support at all levels has been characterized as a major priority and immediate need for the U.S. military services. This paper describes both the conceptual context, based in a flexible group theoretic tabu search (GTTS) framework, and the software implementation of a robust, efficient, and effective prescriptive generalized theater distribution methodology. This methodology evaluates and prescribes the routing and scheduling of multimodal theater transportation assets at the individual asset operational level to provide economically efficient time-definite delivery of cargo to customers.
- **Expanded paper:** Plants take up large amounts of K^+ from the soil solution and distribute it to the cells of all organs, where it fulfills important physiological functions. Transport of K^+ from the soil solution to its final destination is mediated by channels and transporters. To better understand K^+ movements in plants, we intended to characterize the function of the large KT-HAK-KUP family of transporters in rice (*Oryza saliva* cv Nipponbare). By searching in databases and cDNA cloning, we have identified 17 genes (OsHAK1-17) encoding transporters of this family and obtained evidence of the existence of other two genes. Phylogenetic analysis of the encoded transporters reveals a great diversity among them, and three distant transporters, OSHAK1, OSHAK7, and OSHAK10, were expressed in yeast (*Saccharomyces cerevisiae*) and bacterial mutants to determine their functions. The three transporters mediate K^+ influxes or effluxes, depending on the conditions of the experiment. A comparative kinetic analysis of HAK-mediated K^+ influx in yeast and in roots of K^+ -starved rice seedlings demonstrated the involvement of HAK transporters in root K^+ uptake. We discuss that all HAK transporters may mediate K^+ transport, but probably not only in the plasma membrane. Transient expression of the OSHAK10-green fluorescent protein fusion protein in living onion epidermal cells targeted this protein to the tonoplast.

KNOWLEDGE DISCOVERY-2

(EMERGENCY EVACUATION OF BUILDINGS)

- **Original Paper:** Granular materials display a variety of behaviors that are in many ways different from those of other substances. They cannot be easily classified as either solids or liquids. This has prompted the generation of analogies between the physics found in a simple sandpile and that found in complicated microscopic systems, such as flux motion in superconductors or spin glasses. Recently, the unusual behavior of granular systems has led to a number of new theories and to a new era of experimentation on granular systems.
- **Citing Paper:** We review both experimental and theoretical works concerning granular flows. We successively address the regime of slow deformations, which is mainly governed by steric interactions and friction forces, then the rapid flow regime, which deals with inelastic collisions, and lastly the regime of intermittent avalanches.
- **Citing Paper of Citing Paper:** The 2D cellular automata (CA) random model is applied to occupant evacuation considering the influence of human psychology and behavior. Thereby, we put forward some new viewpoints on the performance-based design of building exits: the exit width must be bigger than a certain critical value to ensure a dilute state of evacuation; the optimal value of the exit separation f is independent of the exit width d , but is related to the total width of the building D : f approximate to $0.3D$; a symmetrical layout of the exits should be applied; Bad random factors will occur and lead to a prolonged and fluctuated evacuation time when the value of f is too small.

ILLUSTRATIVE CLIMATE CHANGE EXAMPLE

(GLOBAL WARMING IMPACT ON PARKINSON'S DISEASE)

- **Paper 1:** The projected degrees of theoretical range expansion and increased tick survival by the 2020s, suggest that actual range expansion of *I. scapularis* may be detectable within the next two decades. Seasonal tick activity under climate change scenarios was consistent with maintenance of endemic cycles of the Lyme disease agent in newly established tick populations. The geographic range of *I. scapularis*-borne zoonoses may, therefore, expand significantly northwards as a consequence of climate change this century
- **Paper 2:** Since the discovery of organochlorines, virtually every chemical group of pesticides developed for the control of arthropods is represented among the list of products employed for the control of ticks on cattle. The evolution of tick resistance to acaricides has been a major determinant of the need for new products.
- **Paper 3:** Changes in biochemical status of nerve terminals in the corpus striatum, one of the primary brain regions affected in Parkinson's disease, were studied in groups of C57BL/6 mice treated by ip injection three times over a 2-week period with 3--100 mg/kg heptachlor. On average, the maximal rate of striatal dopamine uptake increased > 2-fold in mice treated at doses of 6 mg/kg heptachlor and 1.7-fold at 12 mg/kg heptachlor. Increases in maximal rate of striatal dopamine uptake were attributed to induction of the dopamine transporter (DAT) and a compensatory response to elevated synaptic levels of dopamine. Significant increase in V_{max} of striatal DAT was not observed at doses > 12 mg/kg, which suggested that toxic effects of heptachlor epoxide may be responsible for loss of maximal dopamine uptake observed at higher doses of heptachlor. In support of this conclusion, polarographic measurements of basal synaptosomal respiration rates from mice treated with doses of heptachlor > 25 mg/kg indicated marked, dose-dependent depression of basal tissue respiration. At doses of 6 and 12 mg/kg heptachlor, which increased expression of striatal DAT, uptake of 5-hydroxytryptamine into cortical synaptosomes was unaffected. Thus, striatal dopaminergic nerve terminals were found to be differentially sensitive to heptachlor. This reduced sensitivity of serotonergic pathways was mirrored in the greater potency of heptachlor epoxide to cause release of dopamine from preloaded striatal synaptosomes in vitro compared to release of serotonin from cortical membranes. These results suggest that heptachlor, and perhaps other organochlorine insecticides, exert selective effects on striatal dopaminergic neurons and may play a role in the etiology of idiopathic Parkinson's disease.

ILLUSTRATIVE VETERANS AFFAIRS EXAMPLE

(*GENISTEIN* FOR TRAUMATIC BRAIN INJURY)

- **Paper 1:** Traumatic brain injury causes a reduction in cerebral blood flow, which may cause additional damage to the brain. The purpose of this study was to examine the role of nitric oxide produced by endothelial nitric oxide synthase (eNOS) in these vascular effects of trauma..... These differences in cerebral hemodynamics between the eNOS-deficient and the wild-type mice suggest an important role for nitric oxide produced by eNOS in the preservation of cerebral blood flow in contused brain following traumatic injury.....
- **Paper 2:** Genistein, a soy phytoestrogen, may improve vascular function, but the mechanism of this effect is unclear. Endothelial-derived nitric oxide (NO) is a key regulator of vascular tone and atherogenesis. Previous studies have established that estrogen can act directly on vascular endothelial cells (EC) to enhance NO synthesis through genomic stimulation of endothelial NO synthase (eNOS) expression. However, it is unknown whether genistein has a similar effect. We therefore investigated whether genistein directly regulates NO synthesis in primary human aortic EC (HAEC) and human umbilical vein EC (HUVEC). Genistein, at physiologically achievable concentrations in individuals consuming soy products, enhanced the expression of eNOS and subsequently elevated NO synthesis in both HAEC and HUVEC, with 1-10 micromol/L genistein inducing the maximal effects. However, the effects of genistein on eNOS and NO were not mediated by activation of estrogen signaling or inhibition of tyrosine kinases, 2 known biological actions of genistein. Genistein (1-10 micromol/L) increased eNOS gene expression (1.8- to 2.6-fold of control) and significantly increased eNOS promoter activity of the human eNOS gene in HAEC and HUVEC, suggesting that genistein activates eNOS transcription. Dietary supplementation of genistein to spontaneously hypertensive rats restored aortic eNOS levels, improved aortic wall thickness, and alleviated hypertension, confirming the biological relevance of the in vitro findings. Our data suggest that genistein has direct genomic effects on the vascular wall that are unrelated to its known actions, leading to increased eNOS expression and NO synthesis, thereby improving hypertension.

ILLUSTRATIVE VETERANS AFFAIRS EXAMPLE (COCOA FOR SPINAL CORD INJURY)

- **Paper 1:** Accumulating evidence over last several years indicates an important role of microglial cells in the pathogenesis of neuropathic pain. Signal transduction in microglia under chronic pain states has begun to be revealed. We will review the evidence that p38 MAPK is activated in spinal microglia after nerve injury and contributes importantly to neuropathic pain development and maintenance. We will discuss the upstream mechanisms causing p38 activation in spinal microglia after nerve injury. We will also discuss the downstream mechanisms by which p38 produces inflammatory mediators. Taken together, current data suggest that p38 plays a critical role in microglial signaling under neuropathic pain conditions and represents a valuable therapeutic target for neuropathic pain management.
- **Paper 2:** Oxidative stress induced by reactive oxygen species has been strongly associated with the pathogenesis of neurodegenerative disorders, including Alzheimer's disease. In this study, we investigated the possible protective effects of a cocoa procyanidin fraction (CPF) and procyanidin B2 (epicatechin-(4beta-8)-epicatechin) - a major polyphenol in cocoa - against apoptosis of PC12 rat pheochromocytoma (PC12) cells induced by hydrogen peroxide (H₂O₂).These results suggest that the protective effects of CPF and procyanidin B2 against H₂O₂-induced apoptosis involve inhibiting the downregulation of Bcl-X(L) and Bcl-2 expression through blocking the activation of JNK and p38 MAPK.

ILLUSTRATIVE BIOSECURITY EXAMPLE

(NANOTECHNOLOGY FOR PLANT VIRUS TRANSFECTION TRACKING)

- **Paper 1:** RNA interference is an exciting field of functional genomics that can silence viral genes. This property of **interfering RNA can be used to combat viral diseases of plants** as well as animals and humans. It is a short sequence of nucleic acid that can bind to the mRNA of the gene and interferes the process of its expression. It is diverse in occurrence as well as in applications. It occurs from nematodes to fungi and can cause gene silencing in plants, animals and human beings. Small interfering RNAs are used to silence plant viral genes and in production of therapeutic drugs against Hepatitis or Immuno-deficiency viruses in human. In this review, we will discuss the history, mechanism and applications of RNA interference in plant, animal and human research.
- **Paper 2:** Gene silencing using short interfering RNA (siRNA) is fast becoming an attractive approach to probe gene function in mammalian cells. Although there have been some success in the delivery of siRNA using various methods, tracking their delivery and monitoring their transfection efficiency prove to be hard without a suitable tracking agent. Therefore, a challenge lies with the **design of an efficient and at the same time, self-tracking, transfection agent for RNA interference. In this paper, chitosan nanoparticles** (NPs) with encapsulated quantum dots (QDs) were synthesized and used to deliver HER2/neu siRNA. Using such a construct, the delivery and transfection of the siRNA can be monitored by the presence of fluorescent QDs in the chitosan NPs. Targeted delivery of HER2 siRNA to HER2-overexpressing SKBR3 breast cancer cells was shown to be specific with chitosan/QD NP surface labeled with HER2 antibody targeting the HER2 receptors on SKBR3 cells. Gene-silencing effects of the conjugated siRNA was also established using the luciferase and HER2 ELISA assays. These self-tracking siRNA delivery NPs will also aid in the monitoring of future gene silencing studies in vivo.

ILLUSTRATIVE BIOSECURITY EXAMPLE

(NANOTECHNOLOGY FOR PLANT VIRUS IMAGING)

- **Paper 1:** The physiological status of plants can nowadays be promptly monitored with non-invasive methods. This opens the possibility to continuously follow-up plant performance and permits to detect stress-induced deviations presymptomatically. Upon stress, plants may synthesize specific compounds, depending on the causal agent. Such compounds may alter the absorption of the light impinging on plant leaves, hence the spectrum of reflected, re-emitted, and transmitted light changes. UV-excited fluorescence imaging specifically allows visualization of the accumulation of phenolic compounds, e.g. those associated with the hypersensitive response to pathogens. By using imaging at regular intervals (time-lapse series) of tobacco mosaic virus (TMV) infection in resistant tobacco we aimed at the description and quantification of the kinetics of blue-green fluorescence compared to the visual development of the disease. Presymptomatic responses to TMV infection were observed with a multicolor fluorescence and reflectance imaging setup. The onset of increases in blue-green and chlorophyll fluorescence were comparable in timing, although further symptom development was strikingly different. Compounds known to accumulate during the hypersensitive response and displaying blue-green fluorescence revealed different dynamics of fluorescence evolution in time. The multichannel imaging system permitted to discern the key components salicylic acid and scopoletin. In contrast, for the compatible interaction between TMV and non-resistant tobacco, no presymptomatic responses were detected on inoculated leaves. This work proves the potential of multispectral imaging to unveil stress-associated signatures, and the power of blue-green fluorescence imaging to monitor accumulation of secondary compounds.
- **Paper 2:** Live cell imaging using CdSe/CdS/ZnS quantum rods (QRs) as targeted optical probes is reported. The QRs, synthesized in organic media using a binary surfactant mixture, were dispersed in aqueous media using mercaptoundecanoic acid (MUA) and lysine. Transferrin (Tf) was linked to the QRs to produce QR-Tf bioconjugates that were used for targeted in vitro delivery to a human cancer cell line. Confocal and two-photon imaging were used to confirm receptor-mediated uptake of QR-Tf conjugates into the HeLa cells, which overexpress the transferrin receptor (TfR). Uptake was not observed with QRs that lacked Tf functionalization or with cells that were presaturated with free Tf and then treated with Tf-functionalized QRs.

ILLUSTRATIVE LOGISTICS EXAMPLE

(OPTIMAL ROUTING LESSONS FROM THE GOLGI COMPLEX)

- **CORE ARTICLE:** The problem of simultaneously allocating customers to depots, finding the delivery routes and determining the vehicle fleet composition is addressed. A multi-level composite heuristic is proposed and two reduction tests are designed to enhance its efficiency. The proposed heuristic is tested on benchmark problems involving up to 360 customers, 2 to 9 depots and 5 different vehicle capacities. When tested on the special case, the multi-depot vehicle routing, our heuristic yields solutions almost as good as those found by the best known heuristics but using only 5 to 10% of their computing time. Encouraging results were also obtained for the case where the vehicles have different capacities. (C) 1997 Elsevier Science B.V.
- **EXPANDED ARTICLE:** We now have considerable understanding of the role of the Golgi complex in the posttranslational modifications of membrane and secretory proteins and of lysosomal hydrolases. It is now also clear that the Golgi plays a key role in the intracellular packaging, addressing, and sorting of these classes of proteins to their final destinations on the secretory and endocytic pathways. While it has been proposed that vesicular budding and fusion underlie entry of proteins into the Golgi from the ER and subsequent movement among its cisternae and exit to their final stations, recent observations indicate that this model may need to be revised based on studies in living cells where vesicular-tubular structures appear to mediate membrane trafficking. This will be a major challenge for investigators in the coming years who will rely again on the use of morphologic techniques of the sort that started it all in 1898. (C) 1998 Elsevier Science B.V. All rights reserved.

BACKUP

RECENT LRDI ACCOMPLISHMENTS - CIRCA 2007

- Literature-based discovery (LBD)
 - Only LBD technique to generate true discovery
 - Conducted LRDI review
 - Showed all claimed discoveries had prior art
 - Showed that our discoveries had no prior art
 - Published findings in open literature
 - *Almost two orders of magnitude more potential discovery than all other researchers combined on benchmark Raynaud's Disease problem*
 - **First demonstration of non-medical application**
 - Journal Special Issue devoted entirely to our discovery
- 58 results (TFSC, February 2008)

RECENT LRDI ACCOMPLISHMENTS - CIRCA 2007 (CONT'D)

- India-China study
 - (influencing India research policy; referenced in intelligence briefings)
- Finland study
 - (ONRG director used to plan Finland trip)
- Nanotechnology
 - (resulted in fifteen publication invitations)

POTENTIAL INITIATIVES

- Characterization
 - Applicable to Each Initiative
 - Identify Infrastructure; Pervasive Themes for Each Topic; E.G.,
 - Electric Power Energy Sources;
 - Water Purification Approaches;
 - Population Control Methods;
 - Space Technology;
 - Environment Damage and Remediation Literature;
 - Transportation Technology;
 - Medical Literature;
 - Learning Approaches;
 - Brain Literature;
 - Agriculture Literature;
 - Etc

POTENTIAL INITIATIVES (CONT'D)

- Discovery
 - New Energy Sources;
 - New Fuels;
 - New Transportation Modes;
 - New Water Purification Approaches;
 - New Disease Treatments;
 - New Education Approaches;
 - New Methods for Discovery;
 - New Cyber-security Defenses;
 - New Data Fusion Approaches;
 - Alternative Network Structures for Infrastructure Survivability

IMPORTANCE TO PLANNING/ MANAGEMENT/ EVALUATION

- PLANNING/ EVALUATION/ MANAGEMENT REQUIRES AWARENESS OF ALL NATIONAL AND GLOBAL S&T
 - S&T COMPLETED
 - S&T ONGOING
 - S&T PLANNED
 - S&T POTENTIAL
- TEXT MINING PROVIDES THIS AWARENESS OF DOCUMENTED S&T AT DIFFERENT TEMPORAL STAGES
- TEXT MINING IS CRITICAL PATH FOR OPTIMAL PERFORMANCE OF NIH MANAGEMENT'S MISSION
 - EXPLOITATION
 - COORDINATION
 - AVOID REDUNDANCY

S&T DEVELOPMENT CYCLE

- PLANNING
- IDENTIFICATION
- SELECTION
- EXECUTION
- TRANSITION

MANAGEMENT DECISION AIDS (S&T DEVELOPMENT CYCLE)

- EACH PHASE REQUIRES MANAGEMENT DECISION
- MANAGEMENT DECISION AIDS (MDAs) HAVE BEEN DEVELOPED TO SUPPORT DECISIONS
 - PEER REVIEW
 - METRICS
 - ROADMAPS
 - TEXT MINING
- ALL MDAs ARE INTER-RELATED
 - E.G., CREDIBLE PEER REVIEW REQUIRES METRICS, ROADMAPS, TEXT MINING

TEXT MINING

- DEFINITION
 - EXTRACTION OF USEFUL INFORMATION FROM TEXT
 - IN MODERN USE, INVOLVES AUTOMATED OR SEMI-AUTOMATED COMPUTERIZED EXTRACTION OF INFORMATION FROM LARGE VOLUMES OF ELECTRONICALLY STORED MATERIAL

TEXT MINING (CONT'D)

- STEPS IN TEXT MINING STUDY
(RETRIEVAL/ PROCESSING/ ANALYSIS)
 - DEFINE OBJECTIVES; DEFINE DATA SOURCES
 - DEVELOP QUERY FOR INFORMATION RETRIEVAL
 - RETRIEVE RECORDS FROM SOURCE DATABASE
 - PROCESS RETRIEVED RECORDS
 - BIBLIOMETRICS
 - COMPUTATIONAL LINGUISTICS
 - PERFORM ANALYSIS/ DRAW CONCLUSIONS₆₆

TEXT MINING (CONT'D)

(ANALYSIS TOOLS)

- EVALUATIVE BIBLIOMETRICS
 - USES COUNTS OF PUBLICATIONS/ PATENTS/ CITATIONS TO DEVELOP S&T PERFORMANCE INDICATORS
 - APPLICATIONS
 - IDENTIFY INFRASTRUCTURE (KEY AUTHORS, CENTERS OF EXCELLENCE) OF TECHNICAL DOMAIN
 - IDENTIFY EXPERTS FOR WORKSHOPS AND PANELS
 - DEVELOP SITE VISITATION STRATEGIES TO ASSESS ORGANIZATIONS GLOBALLY
 - IDENTIFY IMPACTS OF RESEARCH (CITATIONS)

TEXT MINING (CONT'D)

(ANALYSIS TOOLS)

- **COMPUTATIONAL LINGUISTICS**
 - IDENTIFIES TECHNICAL THEMES IN LARGE DATABASES FROM PATTERNS IN TEXT
 - APPLICATIONS
 - ENHANCED INFORMATION RETRIEVAL
 - INCREASED AWARENESS OF GLOBAL TECHNICAL LITERATURE STRUCTURE
 - RADICAL DISCOVERY FROM DISPARATE LITERATURES
 - UNCOVERING UNEXPECTED ASYMMETRIES FROM TECHNICAL LITERATURE
 - ESTIMATING GLOBAL LEVELS OF EFFORT IN S&T SUB-DISCIPLINES
 - TRACKING MYRIAD RESEARCH IMPACTS ACROSS TIME AND APPLICATIONS AREAS

TEXT MINING FOR S&T DEVELOPMENT CYCLE

S&T DEVELOPMENT CYCLE

TEXT MINING CAPABILITY	<u>PLANNING</u>	<u>IDENTIFY</u>	<u>SELECTION</u>	<u>EXECUTION</u>	<u>TRANSITION</u>
INFORMATION RETRIEVAL	IDENTIFY STATE OF ART			IMPROVED RESEARCH	IDENTIFY CUSTOMERS
GLOBAL AWARENESS	EXPLOITATION COORDINATION	EXPLOITATION COORDINATION	PRIORITIZE SELECTION	SUPPORT PROG REVIEW	IDENTIFY CUSTOMERS
DISCOVERY	SET NEW DIRECTIONS	NEW DIRECTIONS	NEW DIRECTIONS	ASSEMBLE TEAMS	
UNEXPECTED ASYMMETRIES	INVESTMENT EMPHASES	INVESTMENT EMPHASES	INVESTMENT EMPHASES	DATA ANALYSES	
ESTIMATE LEVELS OF EFFORT	IDENTIFY GAPS/ DEFICIENCIES	IDENTIFY GAPS/ DEFICIENCIES	PRIORITIZE SELECTION		
TRACK RESEARCH IMPACTS				MARKETING	MARKETING
IDENTIFY INFRASTRUCTURE	CONTACT POINTS	EXPLOITATION COORDINATION	CONTEXT FOR SELECTION		
EXPERTS FOR WORKSHOPS	ADVISORY PANELS	ADVISORY PANELS	EVALUATION PANELS	REVIEW PANELS	TRANSITION PANELS
SITE VISITATION STRATEGIES	POTENTIAL COORDINATION	EXPLOITATION COORDINATION	SUPPORT SELECTION		VISIT 69 POTENTIAL CUSTOMERS

TEXT MINING PILOT PROGRAM THRUSTS

- **FOUR MAJOR THRUST AREAS**
 - LITERATURE-RELATED DISCOVERY
 - COUNTRY ASSESSMENTS
 - SINGLE TECHNOLOGY CORE LITERATURE ASSESSMENTS
 - SINGLE TECHNOLOGY CORE AND EXPANDED LITERATURE ASSESSMENTS

KNOWLEDGE DISCOVERY-1

ASSESSING WEAPONS DEVELOPMENT CAPABILITY

- Identify weapons of interest
- Develop query of weapon component technologies
- Retrieve global component S&T literature
- Retrieve all S&T literature of country of interest
- Link specific documents from global weapon component literature to country literature
 - Latent Semantic Indexing
 - Citation Pathways
- Analyze linkages
- Populate social network with authors of interest
- Identify collaborators

CHINA/USA COMPARISON

- OBJECTIVE: ASSESS CHINA'S S&T PROGRAM RELATIVE TO THAT OF USA
- APPROACH: THREE FOUNDATIONAL S&T ASSESSMENT METRICS, WHETHER FOR A PROJECT, A PROGRAM, OR A NATION'S TOTAL S&T OUTPUT (RNK HANDBOOK OF RIA).
 - RIGHT JOB, JOB RIGHT, AND PRODUCTIVITY/PROGRESS
 - "RIGHT JOB" ADDRESSES THE OVERALL INVESTMENT STRATEGY: ARE THE LARGER S&T OBJECTIVES BEING ADDRESSED CORRECTLY?
 - "JOB RIGHT" ADDRESSES THE S&T APPROACH: ARE THE BEST TECHNIQUES BEING USED TO CONDUCT THE S&T?
 - "PRODUCTIVITY/PROGRESS" ADDRESSES THE S&T OUTPUT AND IMPACT.

CHINA/USA COMPARISON

- “RIGHT JOB” – COMPARED FREQUENCY OF SCI ‘SUBJECT AREA’ CATEGORIES FOR MOST RECENT 100000 ARTICLES FROM CHINA AND USA
 - CHINA HAS STRONG RELATIVE EMPHASIS IN PHYSICAL AND ENGINEERING SCIENCES
 - USA HAS STRONG RELATIVE EMPHASIS IN BIOMEDICAL, SOCIAL, AND PSYCHOLOGICAL SCIENCES

CHINA/USA COMPARISON

- “JOB RIGHT”
- EXAMINED CITATION QUALITY (PERCENT OF CHINA’S PUBLICATIONS IN THE TOP CITATION TIER) OF NANOTECHNOLOGY PUBLICATIONS
 - LOW FOR CHINA, BUT INCREASED FROM 4 PERCENT OF THE US FIGURE IN 1998 TO 20 PERCENT IN 2003
- EXAMINED RELATIVE PUBLICATIONS IN TWO HIGH-QUALITY JOURNALS
 - JOURNAL OF THE AMERICAN CHEMICAL SOCIETY (JACS)
 - JOURNAL OF APPLIED PHYSICS (JAP)
 - OVER THE PAST DECADE, THE CHINA/US RATIO FOR JACS ARTICLES GREW BY AN ORDER OF MAGNITUDE, AND THE RATIO FOR JAP ARTICLES GREW BY MORE THAN A FACTOR OF FIVE.
 - SMALL HIGH-QUALITY COMPONENT IS ACHIEVING RATES OF INCREASE THAT MATCH THE OVERALL GROWTH IN CHINESE⁷⁴ TECHNICAL LITERATURE

CHINA/USA COMPARISON

- EXAMINE PRODUCTIVITY FROM DIFFERENT LEVELS OF AGGREGATION
- CHINA LAGS THE US IN TOTAL SCI PUBLICATIONS BY A FACTOR OF THREE (CIRCA 2009)
- CHINA HAS ESSENTIALLY OBTAINED PARITY WITH THE US IN OVERALL NANOTECHNOLOGY PUBLICATION PRODUCTION (CIRCA 2009)
 - CHINA 20% AHEAD OF USA NANOTECHNOLOGY PUBLICATION PRODUCTION CIRCA 2011
- CHINA IS 60 PERCENT AHEAD OF THE US IN NANOCOMPOSITE PUBLICATION PRODUCTION (CIRCA 2009)
 - 120% AHEAD OF USA CIRCA 2011
- **NEED THIS LEVEL OF DETAIL TO ‘CONNECT THE DOTS’ AND IDENTIFY INVESTMENT STRATEGY**

TEXT MINING EXAMPLES

- LITERATURE-BASED DISCOVERY
- LITERATURE-ASSISTED DISCOVERY
- QUERY DEVELOPMENT
- TAXONOMY – LEVELS OF EMPHASIS
- FACTOR MATRIX – THRUSTS
 - FACTOR MATRIX
 - TAXONOMY
- IMPACT ASSESSMENTS
- UNEXPECTED ASYMMETRIES
- RESEARCH IMPACT-CITATION MINING
- COUNTRY ASSESSMENTS
- GAPS AND DEFICIENCIES
- BIBLIOMETRICS
 - AUTHOR
 - JOURNAL
 - INSTITUTION
 - COUNTRY
 - MOST CITED AUTHORS
 - MOST CITED JOURNALS
 - MOST CITED DOCUMENTS

TEXT MINING EXAMPLES

DISCOVERY - NSF DATABASE STUDY

- OBJECTIVES

- IDENTIFY NSF PROJECTS RELATED DIRECTLY AND INDIRECTLY TO WATER PURIFICATION
- COORDINATION/ JOINT PLANNING/ JOINT FUNDING

- PRODUCTS DESCRIBED

- BAA NOTIFICATION (SPINOFF)

TEXT MINING EXAMPLES

DISCOVERY - NSF DATABASE STUDY (CONT'D)

- BAA NOTIFICATION

- GENERATED EXPANDED LIST OF BAA NOTIFICATION RECIPIENTS
- OBTAINED 300 WHITE PAPERS
- ***(THREE TIMES PREVIOUS YEARS INPUT)***
- APPROX. 2/3 FROM DISPARATE LITERATURES
- TEN TIMES INCREASE SHOULD BE POSSIBLE
 - STARTED LATE IN BAA CYCLE
 - INTERMEDIATE QUERY USED
 - 2.5 WEEKS BEFORE DEADLINE
 - BAA CONTENT NOT INTEGRATED WITH NOTIFICATION

TEXT MINING EXAMPLES

QUERY DEVELOPMENT-NANOTECHNOLOGY

- NANOPARTICLE* OR NANOTUB* OR NANOSTRUCTURE* OR NANOCOMPOSITE* OR NANOWIRE* OR NANOCRYSTAL* OR NANOFIBER* OR NANOFIBRE* OR NANOSPHERE* OR NANOROD* OR NANOTECHNOLOG* OR NANOCLUSTER* OR NANOCAPSULE* OR NANOMATERIAL* OR NANOFABRICAT* OR NANOPOR* OR NANOPARTICULATE* OR NANOPHASE OR NANOPOWDER* OR NANOLITHOGRAPHY OR NANO-PARTICLE* OR NANODEVICE* OR NANODOT* OR NANOINDENT* OR NANOLAYER* OR NANOSCIENCE OR NANOSIZE* OR NANOSCALE* OR ((NM OR NANOMETER* OR NANOMETRE*) AND (SURFACE* OR FILM* OR GRAIN* OR POWDER* OR SILICON OR DEPOSITION OR LAYER* OR DEVICE* OR CLUSTER* OR CRYSTAL* OR MATERIAL* OR ATOMIC FORCE MICROSCOP* OR TRANSMISSION ELECTRON MICROSCOP* OR SCANNING TUNNELING MICROSCOP*)) OR QUANTUM DOT* OR QUANTUM WIRE* OR ((SELF-ASSEMBL* OR SELF-ORGANIZ*) AND (MONOLAYER* OR FILM* OR NANO* OR QUANTUM* OR LAYER* OR MULTILAYER* OR ARRAY*)) OR NANOELECTROSPRAY* OR COULOMB BLOCKADE* OR MOLECULAR WIRE*

TAXONOMY; NANOTECHNOLOGY; SCI

Quantum Phenomena, Optics, Electronics, Magnetism, Tribology, and Films (32983 Rec)	Quantum Phenomena, Optics, Electronics, Magnetism, and Tribology (26077 Rec)	Quantum Phenomena (3326 Rec)	Quantum Dots (2028 Rec)	
			Quantum Wells, Wires, and States (1298 Rec)	
	Films (6906 Rec)	Optics, Electronics, Magnetism, and Tribology (22751 Rec)		Optics and Electronics (16432 Rec)
				Magnetism and Tribology (6319 Rec)
		Thin Films (4760 Rec)		Properties of Thin Films (2251 Rec)
				Applications of Thin Films (2509 Rec)
			Deposition of Films (2146 Rec)	Deposition of Thin Films (1752 Rec)
		Diamond Films (394 Rec)		
Nanotubes, Nanomaterials, Nanoparticles, Polymers, Composites, Metal Complexes, and Bionanotechnology (31742 Rec)	Nanotubes (3211 Rec)	Multi-walled Nanotubes (2350 Rec)	Applications of Carbon Nanotubes (474 Rec)	
			Multi-walled Nanotubes (1876 Rec)	
		Single-walled Nanotubes (861 Rec)		Single- and Double-walled Nanotubes (447 Rec)
			Single-walled Nanotubes (414 Rec)	
	Nanomaterials, Nanoparticles, Polymers, Composites, Metal Complexes, and Bionanotechnology (28531 Rec)	Nanomaterials, Nanoparticles, Polymers, Composites, and Metal Complexes (22686 Rec)		Nanomaterials and Nanoparticles (14263 Rec)
				Polymers, Composites, and Metal Complexes (8423 Rec)
		Bionanotechnology (5845 Rec)		DNA (775 Rec)
			Proteins and Cellular Components (5070 Rec)	

TAXONOMY; NANO INSTRUMENTS; SCI

AFM, NMR, Calorimetry (8423)	NMR, RS, Calorimetry (4684)	NMR, Complexes, Compounds (1546)	NMR Spectroscopy (306)	
			NMR, Complexes, Compounds (1240)	
		RS, Calorimetry (3138)	DSC (1138)	
			Raman Scattering, RS, AFM (2000)	
	AFM (3739)	AFM, Films, Tip, Imaging (2003)		AFM, Film, Tip, Imaging (1055)
				AFM, Film, Substrate, Deposit (948)
		AFM, Films, Deposition, Growth, Substrate (1736)		AFM, Film, Deposit, Substrate, Growth (1511)
				AFM, Magnetic (226)
EM, XRD (19090)	EM (4492)	TEM (2545)	HRTEM (296)	
			TEM (2249)	
		SEM, Films, Composites, Particles, Cells (1947)	SEM, Film, Particle, Cell (1652)	
			SEM, IS (295)	
	XRD, Films (14598)	SEM, XRD, Films, Coatings, Composites (3634)	SEM, XRD (1451)	
			SEM, Film, Coating, Deposit, XRD (2183)	
		XRD, TEM, Thin Films (10964)	TEM, Film, Particle, Nanoparticle, STM (5986)	
			Film, XRD, XPS (4978)	

TEXT MINING EXAMPLES

FACTOR MATRIX – NANOTECHNOLOGY – IDENTIFY THEMES

Cumulative Variance	0.582	1.01
Variance	0.582	0.428
Factor	1	2
z	0.812	-0.119
beta	0.805	-0.111
B	0.75	-0.112
V	0.726	-0.113
C	0.693	-0.13
gamma	0.565	-0.094
alpha	0.521	-0.093
R	0.365	-0.061
crystal structure	0.346	-0.108
XRD	-0.045	-0.29
TEM	-0.049	-0.259
differential	-0.033	-0.253
X-ray diffraction XRD	-0.056	-0.249
x-ray diffraction	-0.006	-0.229
formation	-0.05	-0.226
transmission electron microscopy TEM	-0.053	-0.219
films	-0.078	-0.208
morphology	-0.061	-0.197
electron microscopy SEM	-0.049	-0.192

TEXT MINING EXAMPLES

FACTOR MATRIX TAXONOMY - NANOTECHNOLOGY

- SCI Taxonomy
- Level 1
- Instruments (XRD-TEM-SEM)
- Phenomena/ Properties (Crystal Structure)
- Level 2
- Instruments (XRD-TEM-SEM; Differential Calorimetry)
- Phenomena/ Properties (Crystal Structure; Surface Adsorption [SAM/ Film Deposition])
- Level 3
- Instruments (XRD-TEM-SEM; Differential Calorimetry; AFM)
- Phenomena/ Properties (Crystal Structure; Surface Adsorption [SAM/ Film Deposition]; Photoluminescence [Quantum Dots]; Catalysis

TEXT MINING EXAMPLES

FACTOR MATRIX TAXONOMY - NANOTECHNOLOGY

- These results contrast the differences between the document clustering-based taxonomies and factor matrix-based taxonomies. The document clustering taxonomies are categorized essentially by structures (e.g., nanowires, nanotubes, nanoparticles, films) and phenomena (optics, magnetism). The SCI factor matrix taxonomies are characterized by instruments (XRD, TEM, SEM, AFM, differential calorimetry) and the quantities they measure (crystal structure, surface adsorption, photoluminescence).
- At the first level of the factor matrix taxonomies, the science focus of the SCI, which concentrates on instrumentation and basic scientific phenomena (crystal structure), is clearly seen.
- At the second level, the science focus of the SCI remains the same, with additional instrumentation and measured phenomena shown.

TEXT MINING EXAMPLES

JOURNAL COMPARISONS - CITATIONS

	<i>CORTEX</i>		<i>NEUROPSYCHOLOGIA</i>		<i>BRAIN</i>	
	MOST CITED	LEAST CITED	MOST CITED	LEAST CITED	MOST CITED	LEAST CITED
# AUTH						
Average	3.9	2.8	5.2	2.6	7.1	4.6
Median	4	3	5	1	7.5	4.5
# REFS						
Average	46.3	28	52.5	26.8	68.3	42.4
Median	49	29.5	49	26	62.5	35
# CITES						
Average	21	0.8	71.3	0	166.8	2.8
median	18.5	1	67.5	0	157	3
ORG						
Institution	5	4	2	4	8	2
University	5	6	8	6	2	8
COUNTRY	4 Italy 3 France 1 Austria 1 Belgium 1 Germany	2 Italy 2 USA 2 Germany 2 Japan 1 Neth 1 Australia	4 UK 4 USA 1 Italy 1 Canada	5 USA 2 Italy 1 NZ 1 Neth 1 Australia	5 UK 2 USA 2 Canada 1 Germany	3 Japan 1 USA 1 UK 1 France 1 Italy 1 Canada 1 Germany 1 Neth
TYPE						
Behavior		8		4		
Surgery				1	2	85
Diagnostic-NI		2		5	7	
Diagnostic-INV					1	

TEXT MINING EXAMPLES

JOURNAL COMPARISONS - CITATIONS

- A number of interesting observations may be made from Table 7. First, the most cited articles in *Neuropsychologia* are cited, on average, more than three times as often as the most cited articles in *Cortex*, and the most cited articles in *Brain* are cited, on average, more than twice as often as the most cited articles in *Neuropsychologia*.
- Second, the most cited papers have more authors than the least cited, in all three journals, and the effect is most pronounced in *Neuropsychologia*. Additionally, the average number of authors increases with the average number of citations, ranging from about four authors of the most cited *Cortex* papers to about seven authors of the most cited *Brain* papers.
- Third, the most cited papers have substantially more references than the least cited, in both journals, and the effect is most pronounced in *Neuropsychologia*. Additionally, the average number of citations increases with the average number of references (an effect observed by the first author in recent unpublished text mining studies), ranging from about 46 references in the most cited *Cortex* papers to about 68 references in the most cited *Brain* papers.
- Fourth, there is no clear overall trend in citations as a function of institutional representation. The institution/ (institution + university) ratio (where institution in the table cells should be interpreted as any non-university organization; e.g., research laboratory, clinic, hospital, company) for most cited papers starts at 0.5 for *Cortex*, drops to 0.2 for *Neuropsychologia*, and increases sharply to 0.8 for *Brain*. This ratio for least cited papers starts at 0.4 for both *Cortex* and *Neuropsychologia*, and decreases to 0.2 for *Brain*. Its most dramatic change is from 0.8 for the most cited *Brain* papers to 0.2 for the least cited *Brain* papers.
- Fifth, the most cited papers in *Cortex* are all from continental Western Europe, with heavy representation from Italy and France, while the least cited papers in *Cortex* represent four different continents. The most cited papers in *Neuropsychologia* are, with the exception of Italy, from the UK and North America (with heavy representation from the UK and USA), while the least cited papers have more representation from Western Europe but none from the UK. The most cited papers in *Brain* are from the major English-speaking countries, whereas the least cited are scattered around Western Europe, Asia, and North America.
- Sixth, there is a distinct shift in type of study (the bottom of Table 7) in proceeding from *Cortex* to *Neuropsychologia* to *Brain*. Clinical behavioral studies, many of them essentially case studies, predominate the most cited *Cortex* papers. There are only two papers characterized as Diagnostic-Non-Invasive (e.g., PET, MRI, etc). *Neuropsychologia* has more of a balance between Behavioral and Diagnostic-Non-Invasive in its ten most cited papers. *Brain* shows a heavy emphasis on Diagnostic-Non-Invasive (7/10), two papers on surgical procedures, and one on Diagnostic-Invasive. Based on reading Abstracts from each of these journals, the types as represented in the top ten most cited articles roughly approximate the types of papers published overall. Thus, as citations increase in absolute amounts, the study type transitions from the clinically oriented behavioral focus to the correlates with more objective measurements. Also, as the results from the most cited papers section showed, as the study type transitions from the clinically oriented behavioral focus ('soft' technology) to the more objective measurements ('hard' technology), the most cited papers tend to become more recent.

TEXT MINING EXAMPLES

BILATERAL ASYMMETRY PREDICTION

RATIO OF RIGHT TO LEFT ORGAN CANCER INCIDENCE

MEDLINE –
CASE
STUDIES

ORGAN	RNK	NCI
LUNG	1.358	1.395
KIDNEY	1.024	1.043
TESTE	1.128	1.134
OVARY	1.034	1.038

TEXT MINING EXAMPLES

BILATERAL ASYMMETRY PREDICTION-WRITEUP

- APPROACH

- Four types of cancers were examined: lung, kidney, teste, ovary. For each cancer, Medline case report articles focused solely on 1) cancer of the right organ and 2) cancer of the left organ were retrieved, using information retrieval techniques (5) developed by the author. For example, to obtain the Medline records focused on cancer of the left kidney, the following query was used: (LEFT KIDNEY OR LEFT RENAL) AND KIDNEY NEOPLASMS AND CASE REPORT[MH] NOT (RIGHT KIDNEY OR RIGHT RENAL). The ratio of numbers of right organ to left organ articles was compared to actual patient incidence data obtained from the NCI's SEER database for the period 1979-1998.

- RESULTS

- The results are presented in the table. The first column contains the organ in which the lateral asymmetry is studied, the second column contains the ratio of Medline case report records focused solely on right organ cancer to those focused solely on left organ cancer, and the third column contains a similar ratio obtained from the NCI SEER database of patient incidence records.
- The agreement between the Medline record ratios and the NCI's patient incidence data ratios ranged from within three percent for lung cancer to within one percent for teste and ovary cancer.

TEXT MINING EXAMPLES

CITATION MINING RESULTS – VIBRATING SANDPILES

- Development Category and Cited Paper Theme Alignment of Citing Papers

TECH DEV	33									
TECH DEV	32					1				
TECH DEV	31									
APPL RES	23							1	1	1
APPL RES	22					1			3	
APPL RES	21						1	1		
BAS RES	13	1	2	2	2	2	3		1	
BAS RES	12		2	3	6	4	10	8	10	1
BAS RES	11	3	23	28	27	43	43	30	33	4
		1992	1993	1994	1995	1996	1997	1998	1999	2000
		TIME								
<u>CODE: MATRIX ELEMENT IS NUMBER OF</u>										
<u>PAPERS</u>										

TEXT MINING EXAMPLES

CITATION MINING RESULTS – VIBRATING SANDPILES WRITEUP

- In the figure, the abscissa represents time. The ordinate, in the second column from the left, is a two-character tensor quantity. The first number represents the level of development characterized by the citing paper (1=basic research; 2=applied research; 3=advanced development/ applications), and the second number represents the degree of alignment between the main themes of the citing and cited papers (1=strong alignment; 2=partial alignment; 3=little alignment). Each matrix element represents the number of citing papers in each of the nine categories.
- There are three interesting features on the figure. First, the tail of total annual citation counts is very long, and shows little sign of abating. This is one characteristic feature of a seminal paper.
- Second, the fraction of extra-discipline basic research citing papers to total citing papers ranges from about 15-25% annually, with no latency period evident. This lag-free extra-disciplinary diffusion may have been due to the combination of intrinsic broad-based applicability of the subject matter and publication of the paper in a high-circulation science journal with very broad-based readership.
- Third, a four-year latency period exists prior to the emergence of the higher development category citing papers. This correlates with the results from the bibliometrics component. From the present study, it is not possible to differentiate the reasons for this important result. The latency could have been due to the inability of the technology community to *immediately* recognize the potential applications of the science. Or, it could have been due to the information remaining in the basic research journals, and not reaching the applications community. Or, the time that an application needs to be developed in this discipline is of the order of four years. Thus, the basic science publication feature that may have contributed heavily to extra-discipline citations may also have limited higher development category citations for the latency period.

SELECTED TEXT MINING PUBLICATIONS (LRDI PUBLICATIONS HIGHLIGHTED)

- **Kostoff RN. Literature-related discovery and innovation — update. Technological Forecasting and Social Change (2012). doi:10.1016/j.techfore.2012.02.002.**
- Kostoff RN. China/USA nanotechnology research output comparison—2011 update, Technological Forecasting and Social Change (2012). doi:10.1016/j.techfore.2012.01.007.
- Kostoff RN. Where is the research in the research literature? JASIST. In Press.
- Kostoff RN, Koytcheff RG, and Lau CGY. "Structure of the global nanoscience and nanotechnology research literature". Encyclopedia of Nanoscience and Nanotechnology. American Scientific Publishers. Vol. 23. 417-486. 2011.
- Kostoff RN, Koytcheff RG, and Lau CGY. "Characteristics of the seminal nanotechnology literature". Encyclopedia of Nanoscience and Nanotechnology. American Scientific Publishers. Vol. 12. 271-300. 2011.
- Kostoff RN, Morse SA. "Structure and infrastructure of infectious agent research literature: SARS". Scientometrics. 86:1. 195-209. 2011.
- **Kostoff RN. "Literature-Related Discovery: Potential treatments and preventatives for SARS". Technological Forecasting and Social Change. 78:7. 1164-1173. 2011.**
- Schoeneck DJ, Porter AL, Kostoff RN, Berger EM. "Assessment of Brazil's research literature". TASM. 23:6. 601-621. 2011.
- **Kostoff RN, and Bhattacharya, S. "Identification of militarily-relevant science and technology". Defence Science Journal. 60:3. 259-270. 2010.**
- Kostoff RN. "Expanded information retrieval using full text searching". Journal of Information Science. 36:1. 104-113. 2010.
- Kostoff RN. "The highly cited SARS research literature". Critical Reviews in Microbiology. 36:4. 299-317. 2010.
- Miller R, Kostoff RN. "Assessment of breadth and utility of India's research literature". DTIC Technical Report Number ADA515317. (<http://www.dtic.mil/>) Defense Technical Information Center. Fort Belvoir, VA. 2010.
- **Kostoff RN. "Literature-Related Discovery: Potential treatments and preventatives for SARS". DTIC Technical Report Number ADA 525270. (<http://www.dtic.mil/>) Defense Technical Information Center. Fort Belvoir, VA. 2010.**
- **Kostoff RN. "Literature-related discovery: common factors for Parkinson's Disease and Crohn's Disease". DTIC Technical Report Number ADA525269. (<http://www.dtic.mil/>) Defense Technical Information Center. Fort Belvoir, VA. 2010.**
- Schoeneck DJ, Porter AL, Kostoff RN, Berger EM. "Assessment of Brazil's research literature". DTIC Technical Report Number ADA515318. (<http://www.dtic.mil/>) Defense Technical Information Center. Fort Belvoir, VA. 2010.
- Kostoff RN, Koytcheff RG, and Lau CGY. "Seminal nanotechnology literature: A review". Journal of Nanoscience and Nanotechnology. 9:11. 6239-6270. 2009.
- Chen H, Kostoff RN, Chen C, Zhang J, Vogeley MSE, Börner K, Ma N, Duhon RJ, Zoss A, Srinivasan V, Fox EA, Yang CC, Wei CP. AI and global science and technology assessment. IEEE Intelligent Systems 24:4. 68-88. 2009.
- **Kostoff RN. "A systematic approach to alternative medical procedures". BioScience. 59:9. 734-735. 2009.**
- **Kostoff, R.N. "Literature-Related Discovery: Introduction and Background". Technological Forecasting and Social Change. R.N. Kostoff (ed.). Special Issue on Literature-Related Discovery. 75:2. 165-185. February 2008.**

SELECTED TEXT MINING PUBLICATIONS

(LRDI PUBLICATIONS HIGHLIGHTED)

- Kostoff, R.N., Briggs, M.B., Solka, J.A., Rushenberg, R.L. "Literature-Related Discovery: Methodology". *Technological Forecasting and Social Change*. R.N. Kostoff (ed.). *Special Issue on Literature-Related Discovery*. 75:2. 186-202. February 2008.
- Kostoff, R.N., Block, J.A., Stump, J.A., Johnson, D. "Literature-Related Discovery: Potential Treatments for Raynaud's Phenomenon". *Technological Forecasting and Social Change*. R.N. Kostoff (ed.). *Special Issue on Literature-Related Discovery*. 75:2. 203-214. February 2008.
- Kostoff, R.N. "Literature-Related Discovery: Potential Treatments for Cataracts". *Technological Forecasting and Social Change*. R.N. Kostoff (ed.). *Special Issue on Literature-Related Discovery*. 75:2. 215-225. February 2008.
- Kostoff, R.N., Briggs, M.B. "Literature-Related Discovery: Potential Treatments for Parkinson's Disease". *Technological Forecasting and Social Change*. R.N. Kostoff (ed.). *Special Issue on Literature-Related Discovery*. 75:2. 226-238. February 2008.
- Kostoff, R.N., Briggs, M.B., Lyons, T. "Literature-Related Discovery: Potential Treatments for Multiple Sclerosis". *Technological Forecasting and Social Change*. R.N. Kostoff (ed.). *Special Issue on Literature-Related Discovery*. 75:2. 239-255. February 2008.
- Kostoff, R.N., Solka, J.A., Rushenberg, R.L., Wyatt, J.R. "Literature-Related Discovery: Potential Improvements in Water Purification". *Technological Forecasting and Social Change*. R.N. Kostoff (ed.). *Special Issue on Literature-Related Discovery*. 75:2. 256-275. February 2008.
- Kostoff, R.N., Block, J.A., Solka, J.A., Briggs, M.B., Rushenberg, R.L., Stump, J.A., Johnson, D., Wyatt, J.R. "Literature-Related Discovery: Lessons Learned, and Future Research Directions". *Technological Forecasting and Social Change*. R.N. Kostoff (ed.). *Special Issue on Literature-Related Discovery*. 75:2. 276-299. February 2008.
- Kostoff, R. N., Morse, S., and Oncu, S. "Text Mining of the Anthrax Literature." *Defense Science Journal*. 58:5. 678-685. 2008.
- Kostoff, R.N., Barth, R.B., and Lau, C.G.Y. "Quality vs quantity of publications in nanotechnology field from the Peoples Republic of China". *Chinese Science Bulletin*. 53:8. 1272-1280. April 2008.
- Kostoff, R.N., Barth, R.B., and Lau, C.G.Y. "Relation of seminal nanotechnology document production to total nanotechnology document production - South Korea". *Scientometrics*. 76: 1. 43-67. July 2008.
- Kostoff, R.N., Block, J.A., Solka, J.A., Briggs, M.B., Rushenberg, R.L., Stump, J.A., Johnson, D., Wyatt, J.R. "Literature-Related Discovery". *ARIST*. 43. 243-285. 2008.
- Kostoff, R.N. "Where is the Discovery in Literature-Based Discovery?" in Bruza, P., and Weeber, M., eds. *Literature-Based Discovery*. Springer. Oct 2008.
- Kostoff RN, Koytcheff RG, Lau CGY. "Structure of the nanoscience and nanotechnology applications literature". *Journal of Technology Transfer*. 33:5. 472-484. Oct 2008.
- Kostoff, RN. "Comparison of China/USA science and technology performance". *Journal of Informetrics*. 2:4. 354-363. Oct 2008.
- Kostoff, R.N., Koytcheff, R.G., and Lau, C.G.Y. "Global Nanotechnology Research Metrics". *Scientometrics*. 70:3. 565-601. 2007.
- Kostoff, R.N., Koytcheff, R., and Lau, C.G.Y. "Technical structure of the global nanoscience and nanotechnology literature". *Journal of Nanoparticle Research*. 9:5. 701-724. 2007.
- Kostoff, R.N., Koytcheff, R., and Lau, C.G.Y. "Nanotechnology Instrumentation and its Measurements". *Current Nanoscience*. 3 (2): 135-152 2007.
- Kostoff, R. N., Bhattacharya, S., Pecht, M. "Assessment of China's and India's Science and Technology Literature – Introduction, Background, and Approach". *Technological Forecasting and Social Change*. 74(9). 1519-1538. November 2007.

SELECTED TEXT MINING PUBLICATIONS (LRDI PUBLICATIONS HIGHLIGHTED)

- Kostoff, R. N., Johnson, D., Bowles, C.A., Bhattacharaya, S., Icenhour, A.S., Nikodym, K.F., Barth, R.B., and Dodbele, S. "Assessment of India's research literature". *Technological Forecasting and Social Change*. 74(9). 1574-1608. November 2007.
- Kostoff, R. N., Briggs, M., Rushenberg, R., Bowles, C.A., Icenhour, A.S., Nikodym, K.F., Barth, R.B., and Pecht, M. "Chinese science and technology — Structure and infrastructure". *Technological Forecasting and Social Change*. 74(9). 1539-1573. November 2007.
- Kostoff, R. N., Briggs, M., Rushenberg, R., Johnson, D., Bowles, C.A., Bhattacharaya, S., Icenhour, A.S., Nikodym, K.F., Barth, R.B, Dodbele, S., Pecht, M. "Comparisons of the structure and infrastructure of Chinese and Indian Science and Technology". *Technological Forecasting and Social Change*. 74(9). 1609-1630. November 2007.
- **Kostoff, R.N., Block, J.A., Solka, J.A., Briggs, M.B., Rushenberg, R.L., Stump, J.A., Johnson, D., Wyatt, J.R. "Literature-Related Discovery (LRD)". DTIC Technical Report Number ADA473438. (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2007.**
- Kostoff, R. N., Del Rio, J. A., Cortes, H., Smith, C., Smith, A., Wagner, C.S., Malpohl, G., and Karypis, G. "Clustering Methodologies for Identifying Country Core Competencies". *Journal of Information Science*. 33. 21-40. 2007.
- Kostoff, R.N. "The Difference between Highly and Poorly Cited Medical Articles in the Journal Lancet". *Scientometrics*. 72: 3. 513–520. 2007.
- Kostoff, R.N. and Geisler, E. "The unintended consequences of metrics in technology evaluation". *Journal of Infometrics*. 1(2). 103-114. 2007.
- Kostoff, R.N., Koytcheff, R., and Lau, C.G.Y. "Structure of the Global Nanoscience and Nanotechnology Research Literature". DTIC Technical Report Number ADA461930 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2007.
- Kostoff, R. N. "Normalizing Citation Comparisons for Type of Document." *Scientometrics*.
- Kostoff, R.N., Koytcheff, R., and Lau, C.G.Y. "Global Nanotechnology Research Literature Overview." *Current Science*. 92 (11): 1492-1498. 10 June 2007.
- Kostoff, R.N., Koytcheff, R., and Lau, C.G.Y. "Global Nanotechnology Research Literature". *Science Focus*. 2(2):1-10. 2007.
- Kostoff, R.N., Koytcheff, R., and Lau, C.G.Y. "Applications and Health/ Environmental Impacts of Nanotechnology". *Journal of Technology Transfer*. Online: 11 July 2007.
- Kostoff, R. N., Morse, S., and Oncu, S. "The Seminal Literature of Anthrax Research". *Critical Reviews in Microbiology*. 33: 3. 171 – 181. July 2007.
- Kostoff, R.N. "Validating Discovery in Literature-Based Discovery". *Journal of Biomedical Informatics*. 40: 4. 448-450. 2007.
- Kostoff, R. N. "Text Mining the Biomedical Literature". DTIC Technical Report Number ADA473638. (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2007.
- Kostoff, R. N., Briggs, M., Rushenberg, R., Johnson, D., Bowles, C.A., Bhattacharaya, S., Icenhour, A.S., Nikodym, K.F., Barth, R.B, Dodbele, S., Pecht, M. "An Overview of China's and India's Science and Technology Literature." *Science Focus*. 2(4):1-6. 2007.
- Kostoff, R. N., Briggs, M., Rushenberg, R., Johnson, D., Bowles, C.A., Bhattacharaya, S., Icenhour, A.S., Nikodym, K.F., Barth, R.B, Dodbele, S., Pecht, M. "Assessment of science and technology literature of China and India as reflected in the SCI/SSCI". *Current Science*. 93(8). 1088-1092. 2007.
- Kostoff, R.N., Block, J.A., Solka, J.A., Briggs, M.B., Rushenberg, R.L., Stump, J.A., Johnson, D., Wyatt, J.R. "Literature-related discovery: A review". DTIC Technical Report Number ADA473643. (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2007.
- Kostoff, R.N., Koytcheff, R., and Lau, C.G.Y. "The Growth of Nanotechnology Literature". *Nanotechnology Perceptions*. 2. 229-247. 2006.

SELECTED TEXT MINING PUBLICATIONS (LRDI PUBLICATIONS HIGHLIGHTED)

- Kostoff, R. N., Stump, J.A., Johnson, D., Murday, J., Lau, C., and Tolles, W. "The Structure and Infrastructure of the Global Nanotechnology Literature". *Journal of Nanoparticle Research*. 8 (3-4): 301-321. 2006.
- Kostoff, R. N., Murday, J., Lau, C., and Tolles, W. "The Seminal Literature of Global Nanotechnology Research". *Journal of Nanoparticle Research*. 8 (2): 193-213. 2006.
- **Kostoff, R.N. "Systematic Acceleration of Radical Discovery and Innovation in Science and Technology". *Technological Forecasting and Social Change*. 73 (8): 923-936. 2006.**
- Kostoff, R. N., Johnson, D., Del Rio, J. A., Cortes, H., Bloomfield, L.A., Shlesinger, M. F., and Malpohl, G. "Duplicate Publication and 'Paper Inflation' in the Fractals Literature." *Science and Engineering Ethics*. 12 (3): 543-554. 2006.
- Kostoff, R.N., Rigsby J. T., and Barth, R.B. "Adjacency and Proximity Searching in the Science Citation Index and Google". *Journal of Information Science*. 32:6. 581-587. 2006.
- Kostoff, R.N., and Delafuente, J.C. "The Unknown Impacts of Large Drug Combinations". *Drug Safety*. 29 (3). 183-185. 2006.
- Kostoff, R. N., Johnson D, Del Rio, J. A., Cortes, H., Bloomfield, L.A., Shlesinger, M. F., and Malpohl, G. "Duplicate Publication and 'Paper Inflation' in the Fractals Literature." DTIC Technical Report Number ADA440622 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2006.
- Kostoff, R. N., Tshiteya, R., Bowles, C.A., and Tuunanen, T. "The Structure and Infrastructure of the Finnish Research Literature." *Technology Analysis and Strategic Management*. 18 (2): 187-220. 2006.
- Kostoff, R. N., Tshiteya, R., Bowles, C.A., and Tuunanen, T. "The Structure and Infrastructure of the Finnish Research Literature." DTIC Technical Report Number ADA 442 890. (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2006.
- Kostoff, R.N., Rigsby J. T., and Barth, R.B. "Adjacency and Proximity Searching in the Science Citation Index and Google". DTIC Technical Report Number ADA442888 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2006.
- Kostoff, R. N., Briggs, M., Rushenberg, R., Bowles, C., and Pecht, M. "The Structure and Infrastructure of Chinese Science and Technology." DTIC Technical Report Number ADA443315. (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2006.
- Kostoff, R. N., Johnson, D., Bowles, C.A., and Dodbele, S. "Assessment of India's Research Literature". DTIC Technical Report Number ADA444625 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2006.
- Kostoff, R. N., Morse, S., and Oncu, S. "Text Mining of the Anthrax Literature." DTIC Technical Report Number ADA451571 (<http://www.dtic.mil/>) Defense Technical Information Center. Fort Belvoir, VA. 2006.
- **Kostoff, R.N. "Encouraging Discovery and Innovation". *Science*. 309: 5732. 245-246. 8 July 2005**
- Kostoff, R. N., Karpouzian, G., and Malpohl, G. "Text Mining the Global Abrupt Wing Stall Literature". *Journal of Aircraft*. 42:3. 661-664. 2005.
- Kostoff, R. N., Tshiteya, R., Pfeil, K M., Humenik, J. A., and Karypis, G. "Power Source Roadmaps Using Database Tomography and Bibliometrics". *Energy*. 30:5. 709-730. 2005.
- Kostoff, R. N., and Block, J. A. "Factor Matrix Text Filtering and Clustering." *JASIST*. 56:9. 946-968. July. 2005.
- Kostoff, R. N. "Science and Technology Knowledge Management". in *New Frontiers of Knowledge Management*. (Ed.) Kevin DeSouza. Palgrave Macmillan, United Kingdom. 11-35. 2005.

SELECTED TEXT MINING PUBLICATIONS

(LRDI PUBLICATIONS HIGHLIGHTED)

- Kostoff, R. N., Del Rio, J. A., Smith, C., Smith, A., Wagner, C.S., Malpohl, G., Karypis, G., and Tshiteya, R. "The Structure and Infrastructure of Mexico's Science and Technology". *Technological Forecasting and Social Change*. 72:7. August 2005. Kostoff, R.N., and Shlesinger, M. F. "CAB-Citation-Assisted Background." *Scientometrics*. 62:2. 199-212. 2005.
- Kostoff, R. N. "Exploiting Global Science and Technology". *Marine Corps Gazette*. 89:3. 56-58. March 2005.
- Kostoff, R. N., Buchtel, H., Andrews, J., and Pfeil, K. "The hidden structure of neuropsychology: Text Mining of the Journal Cortex: 1991-2001". *Cortex*. 41:2. 103-115. April 2005.
- Kostoff, R. N. and Martinez, W.L. "Is Citation Normalization Realistic?" *Journal of Information Science*. 31:1. 57-61. 2005.
- Kostoff, R. N. "Systematic Acceleration of Radical Discovery and Innovation in Science and Technology". DTIC Technical Report Number ADA430720 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N. "Science and Technology Metrics". DTIC Technical Report Number ADA432576 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N., Del Rio, J. A., Cortes, H.D., Smith, C., Smith, A., Wagner, C. S., Malpohl, G., and Karypis, G. "Science and Technology Text Mining: Mexico Core Competencies" DTIC Technical Report Number ADA430724 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N., Stump, J.A., Johnson, D., Murday, J., Lau, C., and Tolles, W. "The Structure and Infrastructure of the Global Nanotechnology Literature". DTIC Technical Report Number ADA435984 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N., Murday, J., Lau, C., and Tolles, W. "The Seminal Literature of Global Nanotechnology Research". DTIC Technical Report Number ADA435986 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R.N. and Wasilition, T.P. "Exploiting Global Science and Technology for Army Applications". *Army AL&T Magazine*. Nov-Dec.
- Kostoff, R. N., Tshiteya, R., and Stump, J. A. "Science and Technology Text Mining: Wireless LANs". DTIC Technical Report Number ADA437247 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N., Shlesinger, M., and Tshiteya, R. "Nonlinear Dynamics Roadmaps using Bibliometrics and Database Tomography". *International Journal of Bifurcation and Chaos*. 14:1. 61-92. January 2004.
- Kostoff, R. N., Boylan, R., and Simons, G. R. "Disruptive Technology Roadmaps". *Technology Forecasting and Social Change*. 71:1-2. January-February 2004. 141-159.
- Kostoff, R. N., Shlesinger, M., and Malpohl, G. "Fractals Roadmaps using Bibliometrics and Database Tomography". *Fractals*. 12:1. March 2004. 1-16.
- Kostoff, R.N., Bedford, C.W., Del Rio, J. A., Cortes, H., and Karypis, G. "Macromolecule Mass Spectrometry: Citation Mining of User Documents". *Journal of the American Society for Mass Spectrometry*. 15:3. 281-287. March 2004.
- Kostoff, R. N. "Global Technology Watch". *CHIPS Magazine*. Summer 2004.
- Kostoff, R. N., Block, J. A., Stump, J. A., and Pfeil, K. M. "Information Content in Medline Record Fields". *International Journal of Medical Informatics*. 73:6. 515-527. June.
- Kostoff, R.N. "Scientific Impact of Nations". *The Scientist*. 27 September 2004.

SELECTED TEXT MINING PUBLICATIONS

(LRDI PUBLICATIONS HIGHLIGHTED)

- Kostoff, R. N., Del Rio, J. A., García, E. O., Ramírez, A. M., and Humenik, J. A. "Science and Technology Text Mining: Citation Mining of Dynamic Granular Systems." DTIC Technical Report Number ADA418862 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N., Bedford, C., Del Rio, J. A., Cortes, H., and Karypis, G. "Macromolecule Mass Spectrometry: Citation Mining of User Documents." DTIC Technical Report Number ADA418841 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. "Science and Technology Text Mining: Hypersonic and Supersonic Flow". DTIC Technical Report Number ADA418717 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N., and Geisler, E. "Science and Technology Text Mining : Strategic Management and Implementation in Government Organizations." DTIC Technical Report Number ADA421060 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N., Shlesinger, M., and Tshiteya, R. "Science and Technology Text Mining: Nonlinear Dynamics". DTIC Technical Report Number ADA420998 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N. "Science and Technology Transition Metrics". DTIC Technical Report Number ADA421058 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N., Tshiteya, R., Humenik, J. A., and Pfeil, K M. "Science and Technology Text Mining: Electric Power Sources". DTIC Technical Report Number ADA421789 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N., Andrews, J., Buchtel, H., Pfeil, K., Tshiteya, R., and Humenik, J. A. "Science and Technology Text Mining: Cortex". DTIC Technical Report Number ADA425 056 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N., Block, J. A., Stump, J. A., and Pfeil, K. M. "Information Content in Medline Record Fields". DTIC Technical Report Number ADA423900 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N. and Block, J. A. "Context-Dependent Conflation, Text Filtering and Clustering". DTIC Technical Report Number ADA426072. 1 September 2004 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R.N. "Science and Technology Citation Analysis: Is Citation Normalization Realistic?" DTIC Technical Report Number ADA426271. 8 September 2004 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2005.
- Kostoff, R. N. "A Method for Data and Text Mining and Literature-Based Discovery". Patent Application Publication Number US 2004/0064438 A1. 1 April 2004.
- Kostoff, R. N. "Text Mining for Global Technology Watch". In Encyclopedia of Library and Information Science, Second Edition. Drake, M., Ed. Marcel Dekker, Inc. New York, NY. 2003. Vol. 4. 2789-2799.
- Kostoff, R. N. "Stimulating Innovation". International Handbook of Innovation. Larisa V. Shavinina (ed.). Elsevier Social and Behavioral Sciences, Oxford, UK. 388-400. 2003.
- Kostoff, R. N., Shlesinger, M., and Malpohl, G. "Fractals Roadmaps using Bibliometrics and Database Tomography". SSC San Diego SDONR 477, Space and Naval Warfare Systems Center. San Diego, CA. June 2003.
- Kostoff, R. N., Tshiteya, R., Pfeil, K. M., and Humenik, J. A. "Electrochemical Power: Military Requirements and Literature Structure." Academic and Applied Research in Military Science. 2:1. 5-38. 2003
- Kostoff, R. N. "Data – A Strategic Resource for National Security". Academic and Applied Research in Military Science. 2:1. 169-172. 2003.

SELECTED TEXT MINING PUBLICATIONS

(LRDI PUBLICATIONS HIGHLIGHTED)

- **Kostoff, R. N. "Bilateral Asymmetry Prediction". *Medical Hypotheses*. 61:2. 265-266. August 2003.**
- Kostoff, R.N. "Role of Technical Literature in Science and Technology Development and Exploitation." *Journal of Information Science*. 29:3. 223-228. 2003.
- Hartley, J. and Kostoff, R. N. "How Useful are 'Key Words' in Scientific Journals?" *Journal of Information Science*. 29:5. 433-438. October 2003.
- Kostoff, R. N. "The Practice and Malpractice of Stemming". *JASIST*. 54: 10. June 2003.
- Kostoff, R. N., Karpouzian, G., and Malpohl, G. "Abrupt Wing Stall Roadmaps Using Database Tomography and Bibliometrics". TR NAWCAD PAX/RTR-2003/164 Naval Air Warfare Center, Aircraft Division, Patuxent River, MD. 2003.
- Kostoff, R. N. "Science and Technology Text Mining: Cross-Disciplinary Innovation". DTIC Technical Report Number ADA414807 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 20 June 2003.
- Kostoff, R. N., and DeMarco, R. A. "Science and Technology Text Mining: Analytical Chemistry". DTIC Technical Report Number ADA415945 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2003.
- Kostoff, R. N. "Science and Technology Text Mining: Management Decision Aids". DTIC Technical Report Number ADA415501 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2003.
- Kostoff, R. N., Tshiteya, R., Pfeil, K. M., and Humenik, J. A. "Science and Technology Text Mining: Electrochemical Power." DTIC Technical Report Number ADA415885 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2003.
- Losiewicz, P., Oard, D., and Kostoff, R. N. "Science and Technology Text Mining: Basic Concepts". DTIC Technical Report Number ADA415886 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2003.
- Kostoff, R. N. "Science and Technology Text Mining: Global Technology Watch". DTIC Technical Report Number ADA415863 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2003.
- Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. "Science and Technology Text Mining: Near-Earth Space". DTIC Technical Report Number ADA415928 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2003.
- Kostoff, R. N., Boylan, R., and Simons, G. R. "Disruptive Technology Roadmaps". DTIC Technical Report Number ADA415933 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2003.
- Kostoff, R. N. "Science and Technology Text Mining: Origins of Database Tomography and Multi-Word Clustering". DTIC Technical Report Number ADA416268 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2003.
- Kostoff, R. N., "Science and Technology Text Mining: Comparative Analysis of the Research Impact Assessment Literature and the Journal of the American Chemical Society." DTIC Technical Report Number ADA416267 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2003.
- Kostoff, R. N., and Hartley, J. "Science and Technology Text Mining: Structured Papers". DTIC Technical Report Number ADA417220 (<http://www.dtic.mil/>). Defense Technical Information Center. Fort Belvoir, VA. 2003.
- Kostoff, R. N., Tshiteya, R., Pfeil, K. M., and Humenik, J. A. "Electrochemical Power Source Roadmaps using Bibliometrics and Database Tomography". *Journal of Power Sources*. 110:1. 163-176. 2002.
- Kostoff, R. N., and Hartley J. "Structured Abstracts for Technical Journals". *Journal of Information Science*. 28:3. 257-261. 2002.

SELECTED TEXT MINING PUBLICATIONS

(LRDI PUBLICATIONS HIGHLIGHTED)

- Del Rio, J. A., Kostoff, R. N., Garcia, E. O., Ramirez, A. M., and Humenik, J. A. "Phenomenological Approach to Profile Impact of Scientific Research: Citation Mining." *Advances in Complex Systems*. 5:1. 19-42. 2002.
- Braun, T., Schubert, A., and Kostoff, R. N. "A Chemistry Field in Search of Applications: Statistical Analysis of U. S. Fullerene Patents". *Journal of Chemical Information and Computer Science*. 42:5. 1011-1015. 2002.
- Kostoff, R. N. "Citation Analysis for Research Performer Quality". *Scientometrics*. 53:1. 49-71. 2002.
- Kostoff, R. N. "Biowarfare Agent Prediction". *Homeland Defense Journal*. 1:4. 1-1. 2002.
- Kostoff, R. N. "Overcoming Specialization." *BioScience*. 52:10. 937-941. 2002.
- Kostoff, R. N. "TexTosterone-A Full-Spectrum Text Mining System". Provisional Patent Application. Filed 30 September 2002.
- Kostoff, R. N. "The Extraction of Useful Information from the BioMedical Literature". *Academic Medicine*. 76:12. December 2001.
- Kostoff, R. N., Del Rio, J. A., García, E. O., Ramírez, A. M., and Humenik, J. A. "Citation Mining: Integrating Text Mining and Bibliometrics for Research User Profiling". *JASIST*. 52:13. 1148-1156. 52:13. November 2001.
- Kostoff, R. N., Toothman, D. R., Eberhart, H. J., and Humenik, J. A. "Text Mining Using Database Tomography and Bibliometrics: A Review". *Technology Forecasting and Social Change*. 68:3. November 2001.
- **Kostoff, R. N. "Predicting Biowarfare Agents Takes on Priority". *The Scientist*. 26 November 2001.**
- Kostoff, R. N. "Stimulating Discovery". *Proceedings: Discovery Science Workshop*. November 2001.
- Kostoff, R. N. "Normalization for Citation Analysis". *Cortex*. 37. 604-606. September 2001.
- Kostoff, R. N., and DeMarco, R. A. "Science and Technology Text Mining". *Analytical Chemistry*. 73:13. 370-378A. 1 July 2001.
- Kostoff, R. N. "Intel Gold". *Military Information Technology*. 5:6. July 2001.
- Kostoff, R. N. "Extracting Intel Ore". *Military Information Technology*. 5:5. 24-26. June 2001.
- Kostoff, R. N., and Del Rio, J. A. "Physics Research Impact Assessment". *Physics World*. 14:6. 47-52. June 2001.
- Kostoff, R. N., and Hartley, J. "Structured Abstracts for Technical Journals". *Science*. 11 May. p.292 (5519):1067a. 2001.
- Kostoff, R. N. "The Metrics of Science and Technology". *Scientometrics*. 50:2. 353-361. February 2001.
- Kostoff, R. N., Braun, T., Schubert, A., Toothman, D. R., and Humenik, J. "Fullerene Roadmaps Using Bibliometrics and Database Tomography". *Journal of Chemical Information and Computer Science*. 40:1. 19-39. Jan-Feb 2000.
- Braun, T., Schubert, A. P., and Kostoff, R. N. "Growth and Trends of Fullerene Research as Reflected in its Journal Literature." *Chemical Reviews*. 100:1. 23-27. January 2000.
- Losiewicz, P., Oard, D., and Kostoff, R. N. "Textual Data Mining to Support Science and Technology Management". *Journal of Intelligent Information Systems*. 15. 99-119. 2000.
- Kostoff, R. N., Green, K. A., Toothman, D. R., and Humenik, J. "Database Tomography Applied to an Aircraft Science and Technology Investment Strategy". *Journal of Aircraft*, 37:4. 727-730. July-August 2000.

SELECTED TEXT MINING PUBLICATIONS (LRDI PUBLICATIONS HIGHLIGHTED)

- Kostoff, R. N. "High Quality Information Retrieval for Improving the Conduct and Management of Research and Development". Proceedings: Twelfth International Symposium on Methodologies for Intelligent Systems. 11-14 October 2000.
- Kostoff, R. N. "Implementation of Textual Data Mining in Government Organizations". Proceedings: Federal Data Mining Symposium and Exposition. 28-29 March 2000.
- Kostoff, R. N. "The Underpublishing of Science and Technology Results". The Scientist. 14:9. 6-6. 1 May 2000.
- Kostoff, R. N., Green, K. A., Toothman, D. R., and Humenik, J. A. "Database Tomography Applied to an Aircraft Science and Technology Investment Strategy". TR NAWCAD PAX/RTR-2000/84. Naval Air Warfare Center, Aircraft Division, Patuxent River, MD.
- Del Río, J. A., Kostoff, R. N., García, E. O., Ramírez, A. M., and Humenik, J. A. "Citation Mining Citing Population Profiling using Bibliometrics and Text Mining". Centro de Investigación en Energía, Universidad Nacional Autónoma de México. http://www.cie.unam.mx/W_Reportes.
- Kostoff, R. N. "Science and Technology Text Mining". Keynote presentation/ Proceedings. TTCP/ ITWP Workshop. Farnborough, UK. 12 October 2000.
- Kostoff, R. N. "Implementation of Textual Data Mining in Government Organizations". Proceedings: Federal Data Mining Symposium and Exposition, 28-29 March 2000.