Available online at www.sciencedirect.com
ScienceDirect
Technological Forecasting & Social Change 73 (2006) 923–936
Technological Forecasting and Social Change

Systematic acceleration of radical discovery and innovation in science and technology ☆

Ronald N. Kostoff *

Office of Naval Research, 875 N. Randolph St., Arlington, VA 22217, United States

Received 26 August 2005; accepted 19 September 2005

Abstract

A systematic two-component approach (front-end component, back-end component) to bridging unconnected disciplines and accelerating potentially radical discovery and innovation (based wholly or partially on text mining procedures) is presented. The front-end component has similar objectives to those in the classical literature-based discovery (LBD) approach, although it is different mechanistically and operationally. The front-end component will systematically identify technical disciplines (and their associated leading experts) that are directly or indirectly-related to solving technical problems of high interest. The back-end component is actually a family of back-end techniques, only one of which shares the strictly literature-based analysis of the classical LBD approach. The non-LBD back-end techniques (literature-assisted discovery) make use of the human experts associated with the disparate literatures (disciplines) uncovered in the front-end to generate radical discovery and innovation.

Specifically, in the *literature-assisted discovery* operational mode, these disparate discipline experts could be used as:

1. Recipients of solicitation announcements (BAA, SBIR, MURI, journal Special Issue calls for papers, etc.),
2. Participants in Workshops, Advisory Panels, Review Panels, Roadmaps, and War Games,
3. Points of Contact for Field Science Advisors, Foreign Field Offices, Program Officer site visits, and potential transitions.

☆ DISCLAIMER: The views in this paper are solely those of the author, and do not necessarily represent the views of the Department of the Navy, or any of its components.

* Tel.: +1 703 696 4198; fax: +1 703 696 3098.
E-mail address: kostofr@onr.navy.mil.

0040-1625/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.techfore.2005.09.004

Keywords: Discovery; Innovation; Science and technology; Text mining; Literature-based discovery; Literature-assisted discovery; Radical discovery; Radical innovation; Information retrieval; Unconnected disciplines; Disparate disciplines; Interdisciplinary; Multidisciplinary; Solicitations; Special issues; Workshops; Roadmaps; Advisory panels; Review panels; War games

1. Definitions

Discovery is ascertaining something previously unknown or unrecognized. Innovation reflects the metamorphosis from present practice to some new, hopefully "better" practice. It can be based on existing non-implemented knowledge, discovery of previously unknown information, discovery and synthesis of publicly available knowledge whose independent segments have never been combined, and/or invention. In turn, the invention could derive from logical exploitation of a knowledge base, and/or from spontaneous creativity (e.g., Edisonian discoveries from trial and error) [1].

More generally, radical discovery and radical innovation depend on the source of the inspiration and/or the magnitude of the impact. The more disparate the source of ideas from the target problem discipline, the more radical the potential discovery or innovation. The greater the magnitude of change/impact resulting from the discovery or innovation, the more radical the potential discovery or innovation. The emphasis of the present paper is on the breadth of the source.

2. Background

Literature-based discovery (LBD) is a systematic two-component approach to bridging unconnected disciplines (front-end component, back-end component) based on text mining procedures. LBD allows potentially *radical* discovery and innovation to be hypothesized. Classically, the LBD front-end component has been used to identify the pool of potential discovery and innovation candidates, and the LBD back-end component has been used to hypothesize the potential discovery and innovation *based on literature analysis alone*.

The pioneering LBD study was reported in Swanson's paper hypothesizing treatments for Raynaud's Disease [2]. Subsequent LBD studies were performed by Swanson/Swanson and Smalheiser on other medical problems [e.g., 3–6]. They also developed more formalized analytical techniques for hypothesizing radical discovery [e.g., 7,8]. Other researchers have used variants of Swanson's approach for hypothesizing radical discovery [e.g., 9–12]. Given:

- the length of time since Swanson's pioneering paper (two decades),
- the massive number of medical and technical problems in need of radical discovery, and
- the relatively few articles published in the literature using existing LBD approaches to generate radical discovery (especially articles not published by the Swanson/Smalheiser team and not replicating the initial Raynaud's results),

it is clear that improvements in the fundamental approach and its dissemination and acceptability are required.

Additionally, LBD consists of literature analysis only for the entire process. Yet, the front-end of the process generates the pool of discovery candidates, including all types of documents and their authors, especially from disparate disciplines that historically serve as a powerful source of radical discovery. There appear to be no studies reported in the literature that have made explicit use of the *discipline experts* (identified in the front-end of the LBD process) for generating radical discovery and innovation.

3. Introduction

Discovery and innovation are the cornerstones of frontier research. One of the methods for generating radical discovery and innovation in a target discipline is to use principles and insights from disciplines very disparate to the target discipline, to solve problems in the target discipline. Unfortunately, identifying these linkages between the disparate and target disciplines, and making the subsequent extrapolations have tended to be a very *serendipitous* process. Until now, there has been no fully *systematic* approach to bridging these unconnected target and disparate disciplines. The present paper describes a systematic approach, or more specifically, variants of a systematic approach (based wholly or partially on text mining procedures) for making these connections. One of the virtues of these specific approaches is that most of them can easily be integrated into the operational processes of science and technology (S&T) sponsoring organizations, or research performing organizations. However, some of these approaches do have characteristics of 'disruptive technologies', due to the additional effort required to properly integrate large numbers of concepts representing many disparate disciplines.

There are many examples where acceleration of discovery and/or innovation requires insights and knowledge from 'external' (i.e., indirectly-related or disparate) technical disciplines, sometimes very disparate disciplines. One could envision a solution to 'mine detection' that exploits the remote detection of markers of nitrogen homeostasis in the presence of clinical disorders. In this case, the 'internal' (i.e., core, or directly related) technical discipline is that ordinarily associated with 'mine detection', while the 'external' technical disciplines would be those associated with specific aspects of remote (or possibly in-situ) detection not normally associated with 'mine detection'. The real challenge is to have a systematic process that identifies these 'external' disciplines starting from the 'internal' disciplines (thereby retaining some indirect thread of connectivity between the 'internal' and 'external' disciplines), and then extrapolates the insights and knowledge from these 'external' disciplines to solve problems in the 'internal' discipline or technology of interest.

The challenge has become more critical due to increasing specialization and effective isolation of technical/medical researchers and developers [13]. As research funding and numbers of researchers have increased substantially over the past few decades, the technical literature has increased substantially as a result. Researchers/developers struggle to keep pace with their own disciplines, much less to develop awareness of other disciplines. Thus, we have the paradox that the *expansion of research* has led to the *balkanization of research*! The resulting balkanization serves as a barrier to cross-discipline knowledge transfers, and retards the progress of discovery and innovation [13].

There are two main text mining avenues for extrapolating knowledge and insights from one discipline/technology to another: *literature-based discovery* and *literature-assisted discovery*. The *literature-based discovery* approach uses technical experts to access and examine the literature from 'external' disciplines to help solve problems in the 'internal' discipline. The *literature-assisted discovery* approach uses technical experts from 'external' disciplines in a variety of interactive and/or independent creative modes for the same purpose.

The main thesis of this paper is that the scientific community has not made adequate use of these 'external' discipline sources of knowledge to accelerate potentially radical discovery and innovation. Further, very substantive quality enhancements to funding agency S&T programs, individual research projects, journal Special Issues, and multi-disciplinary teams and organizations are possible at relatively small marginal costs, if we can *systematically* improve access to the limitless sources of 'external' discipline/technology information.

4. Radical discovery and innovation concept

Text mining is the extraction of useful information from large volumes of text [14,15]. The author's text mining effort over the past decade has been developing methods to systematically access external sources of information that could contribute to problem solving for specific technical disciplines, technologies, systems, operations, or technical problems in general [e.g., 16–25]. These methods have been integrated to form the following systematic approach for accelerating radical discovery and innovation.

Fig. 1 contains a schematic of our text mining approach to discovery. The inner circle represents the core literature of the problem to be solved. In the example for Fig. 1, the problem to be solved is identifying 'improved' alternatives to existing water purification technologies, where 'improved' could encompass any combination of lower cost, lower energy use, lower maintenance, higher reliability, lighter weight, and improved modularity for field assembly. Thus, the core literature is the existing (more or less commonly accepted) water purification literature. The annular region between the inner and outer circles represents literatures related more indirectly to the core literature.

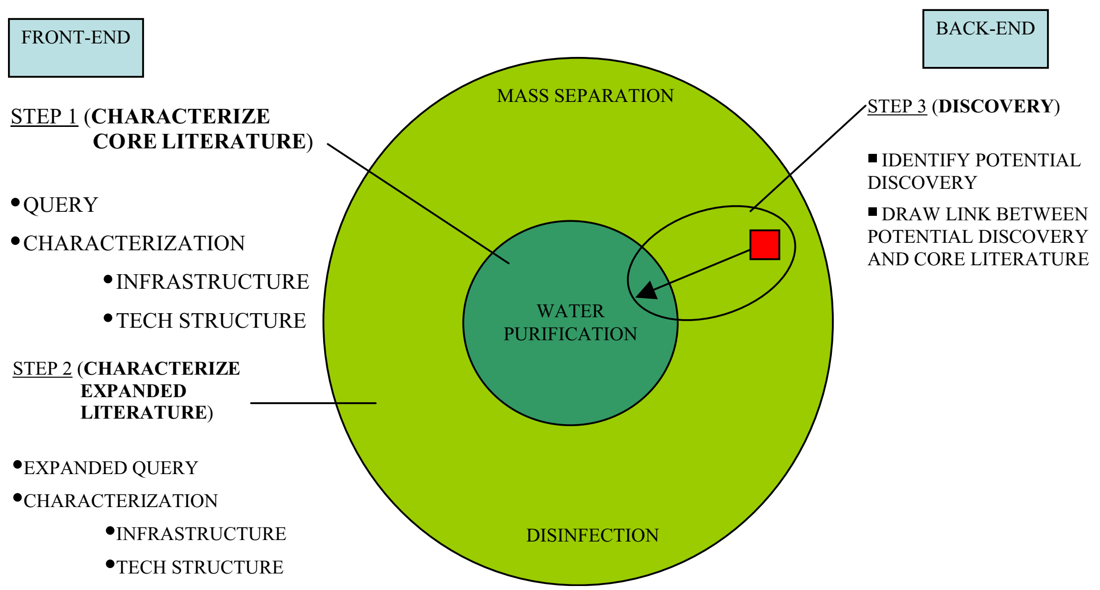

Fig. 1. The discovery process presented in the present paper is divided into two components, a front-end and a back-end.

5. Front-end

5.1. Step 1

The front-end component (summarized to the left of the figure) contains two major steps: characterization of the core literature (Step 1), and characterization of the expanded literature, including identification of technical experts associated with this literature (Step 2). In Step 1, a query to retrieve the core literature is developed iteratively [e.g., 26,27]. Once the core literature has been retrieved with this query, it is subject to text mining [14–25]. Bibliometrics provides the technical infrastructure (key authors/institutions/countries/journals, etc.) of the core literature [e.g., 16–25], and computational linguistics provides the technical structure (technical thrusts, hierarchical taxonomies) of the core literature [e.g., 16–25]. Step 1 reflects the scope of many of our mono-technology text mining studies to date [e.g., 16–25].

The criticality of Step 1 cannot be overemphasized. The core literature represents the starting point for the expansion processes. The derived expanded literature determines the pool of discovery candidates. Any gaps in the core literature will be reflected as gaps in the pool of potential discovery candidates. Therefore, it is imperative that the core literature be as complete and comprehensive as possible for the discovery application.

References [26,27] above summarize the author's approach to query generation for core literature retrieval. Extensive exploitation of co-occurrence phenoma across many attributes is made. For discovery purposes in particular, techniques that specifically exploit the underlying semantic structure of the core literature should also be used, in addition to strictly co-occurrence techniques. The author has made extensive use of factor analysis in understanding the semantic/conceptual structure of retrieved literatures, and has made less formal use of factor analysis for query refinement of the core literature. The factor matrix filtering technique [28] was developed to exploit the underlying semantic structure of a retrieved literature for the purpose of identifying high technical content phrases based on the strength of their contribution to semantic concepts. This is another approach for selecting new query terms.

More formal techniques that exploit the semantic structure, such as latent semantic indexing [10,29], should also be examined for core literature definition. Whether these semantic structure exploitation techniques offer more than properly conducted attribute co-occurrence techniques [e.g., 27] remains to be demonstrated in practice. A multitude of information retrieval techniques have been examined for more than a decade at the TREC conferences [http://trec.nist.gov/pubs.html], and the interested reader is advised to examine the proceedings of these conferences.

5.2. Step 2

5.2.1. Overview

In Step 2, the query developed in Step 1 is generalized and expanded, again iteratively (see the next section for some practical techniques to generate the expanded query). This expanded query will retrieve records from literatures more indirectly related to the core literature. Insights and principles from these disparate literatures/technical disciplines can be extrapolated to solve problems of the core literature. Thus, in the example on Fig. 1, the core water purification literature query is expanded to cover/retrieve all of mass separation and disinfection documents. Insights from very disparate mass separation and disinfection approaches can then be extrapolated to solve problems in water purification.

5.2.2. Details of query expansion

This section provides more detail on query expansion for discovery and innovation. The objective is to generalize the query terms while maintaining a delicate balance: the generalized terms bear some relation to the initial core literature retrieval terms while the relation is sufficiently indirect for the two literatures to be considered disjoint. Also, the expanded query terms are not overly general such that an unwieldy amount of data is retrieved, or the records retrieved are so distant from those of the core literature that impacts will be minimal [30].

The approach consists of examining each core literature query term, and testing different levels of generalizing the term to insure that the above objectives are met. If the records retrieved from the previous iteration are clustered, then terms for the expanded query should be selected such that each main theme from the clustering is adequately represented in the query [31]. Each term being considered for the expanded query should be tested in the source database (e.g., Science Citation Index [SCI]) for retrieval efficiency of relevant records. In some cases, very general forms of the term should be inserted in the source database, and the retrieved records analyzed for more specific variants of the overly general form of the term.

At this point, an example may be illuminating. Consider the topics represented in Fig. 1. The first precursor discovery objective is to expand the core water purification literature to include a more general indirectly-related literature containing potential discovery and innovation candidates. Clustering of the core literature may show very narrow and specific aspects of the assumed two main themes: distillation, membrane filtering *(mass separation)*, ozonation, chlorination *(disinfection)*. Phrases selected for the query should be drawn from each of the more generic versions of the main themes, shown italicized in the previous sentence.

For example, suppose WATER PURIFICATION is one of the query terms for retrieving the core literature. How could it be generalized for expansion, according to the principles set forth above? One approach is incremental generalization. WATER is a sub-set of LIQUID. Therefore, WATER PURIFICATION could be generalized incrementally to LIQUID PURIFICATION. Use of this query term in the source literature would retrieve documents on purification of liquids in addition to water, and any novel concepts used to purify liquids other than water could be extrapolated to help solve the water purification problem.

Before adding this more general term to the query, LIQUID PURIFICATION should be inserted into the SCI search engine, and the retrieval sampled to insure that a high fraction of records relevant to mass separation are being retrieved. In turn, LIQUID is a sub-set of FLUID. Therefore, LIQUID PURIFICATION could be generalized incrementally to FLUID PURIFICATION, and now concepts from the additional gas purification documents could be extrapolated to water purification improvements. The same following steps above would be repeated. The next generalization might be to MASS PURIFICATION, and so on.

How broadly should a query term be generalized? The more directly related the expanded literature is to the core literature, the more obvious will be the connections, but the lower will be the probability for radical discovery and innovation. The more indirectly related the expanded literature is to the core literature, the less obvious will be the connections, but the higher will be the probability for radical discovery and innovation. Thus, if radical discovery and innovation are the goals, the broadest expansion consistent with available resources and reasonable numbers of links in the relational chain should be utilized.

For the WATER PURIFICATION query term being discussed, a second approach to generalization for expansion is to filter the core literature records for all phrases containing the word PURIFICATION.

Then, each reasonable query expansion term candidate based on PURIFICATION can be checked following the steps above.

A third approach to generalization is to select one of the phrase words that is too generic to use as a stand-alone query term (e.g., WATER or PURIFICATION) for further examination. Since purification is the technology of interest, the word PURIFICATION would be inserted into the SCI search engine, and thousands of records retrieved. Text analyses would be performed on these retrieved records, and all phrases containing the word PURIFICATION would be extracted, and examined. Each PURIFICATION phrase variant proposed for the query would follow the same checks described above. While time consuming, this is the author's preferred approach for examining foundational terms for query expansion. In a water purification study, for example, this approach could be used for the very generic terms of foundational importance to the core separation processes (e.g., REMOVAL, SEPARATION, PURIFICATION, EXTRACTION, etc.). This approach can provide quite comprehensive query terms. Because of the time involved for this latter expansion approach, only the most important generic roots should be examined. All other terms could be examined for expansion using the first or second methods described. As suggested in the core literature development section, once the expanded literature has been generated using the approaches of the preceding paragraph, it could be broadened further using techniques that exploit the underlying semantic structure [e.g., 10,29].

How large should resultant queries be? For queries whose objective is retrieval of a statistically representative sample of documents from the source literature to define the core literature, our marginal utility approach can be used to determine cutoff [22]. Basically, more query terms provide increased refinement of a fixed scope literature (e.g., the existing water purification literature). In the reference example [22] (a text mining analysis of the NonLinear Dynamics literature), a 156 term query was reduced to 100 terms, with essentially no loss in fidelity of retrieval of the NonLinear Dynamics literature. Other queries used in the past for this objective ranged from a handful of terms to hundreds of terms, depending strongly on the technology being examined.

For queries whose objective is retrieval of potential discovery items, most comprehensive retrieval coverage of an expanding scope literature, consistent with high retrieval precision (mainly relevant records), is required. The numbers of query terms will be higher than in the first case, and queries up to many hundreds of terms in length (depending on the specific technologies being studied) are possible, and in fact have been generated.

6. Back-end

The back-end component contains the discovery step, which itself contains two sub-components. The first sub-component is identification of potential discovery and innovation candidates from the expanded literature, and the second sub-component is drawing the linkages between the potential discovery/innovation candidates and the core literature. As will be shown in this section, there are many ways to identify potential discovery and innovation candidates, and to draw the subsequent linkages. These techniques differ mainly by the approach mechanics and the types of people used to identify the discovery and innovation candidates. The two main discovery and innovation approach types (Literature-Based, Literature-Assisted) are described now.

6.1. Literature-based discovery

We can use systematic techniques to identify potential discovery and innovation based strictly on the 'external' literature, an approach known as literature-based discovery (LBD) [2,8,9]. LBD is useful in the planning and concept identification phases of the S&T development cycle. The literature-based approach can be viewed as a very sophisticated type of literature survey, and represents a somewhat different way of doing business for most S&T sponsoring agencies, researchers, and technical journals. *Done properly, LBD has the potential of generating at least an order of magnitude more true discovery than what has been reported in the LBD literature (as we are in the process of demonstrating).*

6.2. Literature-assisted discovery

We can identify technical experts associated with the 'external' indirectly-related disciplines, and then have them focus their expertise on solving problems of interest from the 'internal' disciplines. This literature-assisted people-based approach could easily be incorporated into most S&T sponsoring agencies' existing operational procedures. However, in some applications, proper handling of the infusion of large numbers of concepts and insights from disparate disciplines will acquire the characteristics of 'disruptive technologies'.

Thus, the differences between paths A and B above are in the 'back-end', in (1) how the linkages between the 'external' and 'internal' disciplines are made, and (2) who makes the linkages.

The ultimate goal should be incorporation of both approaches in parallel, to exploit the strengths of each approach while eliminating the weaknesses. This synergy would provide the *comprehensiveness and objectivity* of the completely literature-based approach coupled with the *interaction and feedback* of the literature-assisted people-based approach [1].

With respect to the literature-assisted approach, how would the 'external' discipline experts be incorporated into different components of the overall research enterprise's operations, for the purpose of enhancing and accelerating discovery and innovation? The following literature-assisted *options* are a sample of what is possible.

6.2.1. Solicitations—science and technology sponsoring organizations

Government agencies and private foundations generate numerous solicitations for proposals and/or new ideas for solving problems. In the United States, these include Broad Agency Announcements (BAAs), Small Business Innovation Research (SBIR) solicitations, and other similar types. For the Federal agencies, these solicitations are usually advertised in some widely available forum (e.g., FedBizOpps) and, in parallel, some announcements of the solicitation are disseminated to technical experts deemed knowledgeable about the topic. *These targeted announcements are extremely important, since they insure that the recipient will be aware of the solicitation*. The numbers of announcements are usually modest, because of limited prior access to potentially relevant communities.

OPTION 1. The *first option* is that the 'external' discipline experts identified through the text mining, as well as the comprehensive list of 'internal' discipline experts, constitute the bulk of the announcement distribution list. In this way, their expertise in a directly or indirectly-related discipline could be brought to bear on solving the sponsor's problem of interest, with their motivation amplified by the potential for funding, if successful. These 'internal' and 'external' discipline experts would also serve as the gateway to identifying additional technical experts (in their specific disciplines) not associated with the particular literatures accessed (e.g., through common professional societies, institution sub-divisions, attendance at conferences with the experts identified through the strictly literature approach), and thereby adding these additional technical experts to the problem-solving process. Use of the expanded notification list could result in an order of magnitude more proposals, and perhaps two orders of magnitude more potentially radical discovery and innovation proposals.

Potential consequences ('side-effects') resulting from this new approach to solicitations include (1) *substantial* increases in numbers of proposals (*as we have demonstrated successfully*), (2) need to expand diversity of reviewers' technical disciplines to insure interdisciplinary proposals receive balanced evaluation [13], and (3) need to facilitate/stimulate discovery process by improved notification instructions. Consequences (1) and (2) could have modest 'disruptive technology' characteristics, especially if very large numbers of proposals from experts in many disparate disciplines are received.

6.2.2. Solicitations—science and technology journals

Many technical specialty journals are structured on centuries-old research archival and dissemination models. Their scope is 'stove-piped' about quite narrow themes. This parochialism is further compounded by the increasing influence of Impact Factor as a publication metric target, restricting the types of articles published to increasingly narrower bands. The publication trend is toward more narrowly discipline-focused articles that will receive high citations. The trend is away from articles that encompass very diverse disciplines, may not receive high citations on average due to their interdisciplinary nature [13], but could stimulate more radical discovery and innovation.

My recent citation studies of specific technical journals and of specific multi-journal technical disciplines show graphically the narrow dispersion of highly cited paper types, especially in the technical specialty journals. This trend needs to be reversed if the technical journals are to assume their rightful positions as engines of innovation. The following paragraphs offer one approach for reversing this trend.

Most technical journals produce Special Issues periodically. These Special Issues tend to focus on a single topic, and usually have recognized experts on the specific topic present their perspectives. The papers tend to focus on comprehensiveness of coverage about the specific topic, rather then venturing into very disparate disciplines in a search for discovery.

There are two main avenues by which Special Issue papers are solicited. One is the Guest Editor (usually a recognized expert in the Special Issue topic) inviting other recognized experts known to him/her. The second is the Guest Editor/journal placing 'call for papers' ads in prior issues of the journal, or other closely-related topic-centric journals. In both cases, the result is the same: papers centered closely about the topic of interest.

However, the Special Issue concept could be expanded to emphasize radical discovery and innovation. As in the science and technology sponsoring organization example, notification of the projected Special Issue could be sent to the technical experts identified in the front-end of the text mining discovery study. These experts would be encouraged to submit papers for the Special Issue that involved extrapolation of insights and principles from their own technical specialties to solving problems in the Special Issue topic. In this operational mode, the technical journals would serve proactively as the engines for radical discovery and innovation.

OPTION 2. The *second option* is that the 'external' discipline experts identified through the text mining, as well as the comprehensive list of 'internal' discipline experts, constitute the bulk of the journal Special Issue announcement distribution list. In this way, their expertise in a directly or indirectly-related discipline could be brought to bear on solving the Special Issue's particular problem of interest, with their motivation amplified by the potential for journal publication, if successful.

The potential consequences from this new mode of Special Issue operation mirror those of the first option: more papers than normal, greater topical diversity in papers, and need to facilitate discovery and innovation. It is strongly recommended that the Special Issue be expanded in size from the normal journal practice, to accommodate the expected increase in number of submittals. Sponsorship of such Special Issues by the science and technology funding organizations seems appropriate. Modest honoraria could also be provided to authors as further motivation for participation.

6.2.3. Advisory panels

Government agencies and private organizations convene numerous advisory panels or groups of independent advisors for the purpose of providing expert technical advice on problems of present and future interest. Unfortunately, many of these advisory groups are somewhat parochial, both in terms of technical scope and people, thereby limiting the breadth of their recommendations.

OPTION 3. The *third option* is that the 'external' (and 'internal') discipline experts identified through the text mining constitute a significant portion of the members of these groups.

6.2.4. Workshops

Government agencies and private organizations conduct numerous workshops for the purpose of generating new project ideas and directions. Many of these workshop participants are somewhat parochial, both in terms of technical scope and people. Additionally, venture capital organizations, and other components of Wall Street, have great needs for workshops that could provide insight on the potential of emerging technologies. It would be useful for these organizations to know whether the core technology of interest is amenable to improvement, and whether some of the indirectly-related technologies can offer potential solutions across many different core technologies.

OPTION 4. The *fourth option* is that the 'external' (and 'internal') discipline experts identified through the text mining constitute a significant portion of the participants at these workshops [1]. If the workshops are conducted in tandem with the solicitation processes above, then the results from the solicitations could be used to narrow the pool of candidates for these workshops. The workshop attendees could be drawn from the solicitation announcement recipient group that submitted proposals to the sponsoring organization solicitations, or the solicitation announcement recipient group that submitted papers to the technical journal solicitations. At a minimum, the members of these groups had sufficient insight to perceive how concepts from their areas of expertise could be extrapolated to solve problems of interest in the core technology.

6.2.5. Review panels

Government agencies and private organizations conduct numerous review panels during the execution of their S&T programs. Some agencies use such panels to help evaluate proposals and provide recommendations. Many of these review/evaluation panels are somewhat parochial, both in terms of technical scope and people. This limits discussion of the breadth of approaches that could be used to achieve the program objectives.

OPTION 5. The *fifth option* is that a portion of the reviewers be drawn from the 'external' (and 'internal') discipline experts identified through the text mining.

6.2.6. Roadmaps

Some government agency and private organization programs generate technology roadmaps (see Ref. [32] for a description of technology roadmaps) as part of their planning processes, and/or as part of their review processes. Many of these roadmap development teams have limited perspectives, both in terms of technical scope and people. The breadth of these roadmaps is limited by the breadth of their developers.

OPTION 6. The *sixth option* is that a sub-set of the 'external' (and 'internal') discipline experts identified through the text mining constitute a significant portion of the roadmap development team. As in the workshop option, the solicitation step could be used to filter the candidate pool for prospective roadmap development team members.

6.2.7. Points of contact

Many government agency components require points of contact (POCs) for obtaining information to solve problems. These include the Field Science Advisors for military organizations, Foreign Field Offices of government agencies, Program Officers for site visits, and Program Officers to identify potential transitions. In practice, many of these POCs accessed derive from limited personal knowledge, both in terms of technical scope and people. The breadth of the information obtained from the POCs is limited by their breadth of expertise.

OPTION 7. The *seventh option* is that the 'external' (and 'internal') discipline experts identified through the text mining constitute a very significant portion of the POCs accessed in practice. Again, as in the workshop option, the solicitation step could be used to filter the candidate pool for prospective POCs.

6.2.8. Organization and team structuring

Technical teams and organizations can be structured to maximize the potential for radical discovery and innovation, based on the principles presented in the paragraphs above. Multi-disciplinary groups/structures such as Integrated Product Teams (IPTs), Multi-Disciplinary Research Programs of the URI (MURIs), and Cooperative Research and Development Agreements (CRADAs) could be assembled based on the technical disciplines and technical experts identified at the front-end of the discovery process above. Organizations such as Centers of Excellence with a defined core competency, and large laboratories with multiple core competencies, could be structured to incorporate the technical disciplines surrounding their core identified at the front-end of the discovery process.

OPTION 8. The *eighth option* is that the 'external' (and 'internal') disciplines identified through the text mining constitute a significant portion of the teams and organizations. Where possible, the solicitation step could be used to filter the candidate pool for prospective team and organization members.

6.2.9. War games

War games model and/or simulate different pre-combat, combat, and post-combat scenarios. Each Title 10 war game, in addition to major and minor players, the assessors, and the game controlling authority, also has various supporting cells to provide expertise and insight in key areas. For example, the Science and Technology Cell's primary function at the *Global 98* game was to provide future projections of technology that could be available to the game players, and to suggest applications resulting from ongoing R&D. The effectiveness of the advice from such S&T cells depends on the breadth of technical understanding of the cells' members. Additionally, the value of the war games depends in part on the technology capabilities designed into the games and the post-games evaluation of the role that technology played in impacting tactics and strategy.

OPTION 9. The *ninth option* is that the 'external' (and 'internal') disciplines identified through the text mining constitute a portion of the supporting S&T cells' memberships, the membership of the games' technology designers, and the membership of the post-games' technology evaluators.

7. Summary and conclusions

I have proposed the identification and exploitation of diverse literatures, and their representative experts, to help solve problems of interest through potentially radical discovery and innovation. The approach is based on our demonstrated text mining techniques. The essential element is development of comprehensive and precise queries for retrieving the expanded literature of potential discovery candidates, followed by exploitation of these retrieved literatures and their associated technical expert representatives.

I have identified a number of pathways by which these literatures and people could be integrated with present business practices, or could be integrated with slight modifications if desired. *This group of 'external' literature accession techniques has the highest benefit/cost ratio of any techniques I know for enhancing and accelerating radical discovery and innovation*.

Additionally, the "structural holes" research of Professor Burt [33], which identifies 'structural holes' as the weakly-linked or non-linked region between different technical disciplines, has shown that most innovations studied have been the result of drawing on insights from unconnected, sometimes very disparate, disciplines/technologies. These findings support the thesis that is the basis for my proposal above. Translated into practice, the following important guidelines can be drawn.

1. If we are interested in meeting short-term deadlines efficiently with specific well-defined technology products, then homogeneous well-coordinated and long-standing groups are most useful.
2. If we are interested in radical discovery and innovation, including innovation in the advanced technology demonstration of system integration sense, then we need to incorporate people *we don't know* representing disciplines with which *we are not familiar*. It is difficult to get '*out-of-the-box*' thinking from people who have spent their careers '*in-the-box*'!

The generic literature-based and literature-assisted discovery approaches provide a systematic and objective guide to selecting the most appropriate disciplines and people for accelerating potentially radical discovery and innovation.

References

[1] R.N. Kostoff, Stimulating innovation, in: Larisa V. Shavinina (Ed.), International Handbook of Innovation, Elsevier Social and Behavioral Sciences, Oxford, UK, 2003, pp. 388–400.
[2] D.R. Swansonm, Fish oil, Raynauds syndrome, and undiscovered public knowledge, Perspect. Biol. Med. 30 (1) (1986 (Fall)) 7–18.
[3] D.R. Swanson, Migraine and magnesium-11 neglected connections, Perspect. Biol. Med. 31 (4) (1988 (Summer)) 526–557.

[4] D.R. Swanson, Somatomedin-C and arginine—implicit connections between mutually isolated literatures, Perspect. Biol. Med. 33 (2) (1990 (Winter)) 157–186.

[5] N.R. Smalheiser, D.R. Swanson, Assessing a gap in the biomedical literature—magnesium-deficiency and neurologic disease, Neurosci. Res. Commun. 15 (1) (1994 (July–August)) 1–9.

[6] D.R. Swanson, N.R. Smalheiser, A. Bookstein, Information discovery from complementary literatures: categorizing viruses as potential weapons, J. Am. Soc. Inf. Sci. Technol. 52 (10) (2001 (August)) 797–812.

[7] D.R. Swanson, N.R. Smalheiser, An interactive system for finding complementary literatures: a stimulus to scientific discovery, Artif. Intell. 91 (2) (1997).

[8] N.R. Smalheiser, D.R. Swanson, Using ARROWSMITH: a computer-assisted approach to formulating and assessing scientific hypotheses, Comput. Methods Programs Biomed. 57 (3) (1998 (Nov)) 149–153.

[9] M.D. Gordon, R.K. Lindsay, Toward discovery support systems: a replication, re-examination, and extension of Swanson's work on literature-based discovery of a connection between Raynaud's and fish oil, J. Am. Soc. Inf. Sci. 47 (2) (1996 (Feb)) 116–128.

[10] M.D. Gordon, S. Dumais, Using latent semantic indexing for literature based discovery, J. Am. Soc. Inf. Sci. 49 (8) (1998 (June)) 674–685.

[11] M. Weeber, H. Klein, L.T.W. de Jong-van den Berg, et al., Using concepts in literature-based discovery: simulating Swanson's Raynaud-fish oil and migraine–magnesium discoveries, J. Am. Soc. Inf. Sci. Technol. 52 (7) (2001 (May)) 548–557.

[12] J. Stegmann, G. Grohmann, Hypothesis generation guided by co-word clustering, Scientometrics 56 (1) (2003) 111–135.

[13] R.N. Kostoff, Overcoming specialization, BioScience 52 (10) (2002) 937–941.

[14] R.N. Kostoff, Text mining for global technology watch, in: M. Drake (Ed.), Second edition, Encyclopedia of Library and Information Science, vol. 4, Marcel Dekker, Inc., New York, NY, 2003, pp. 2789–2799.

[15] M. Hearst, Untangling text data mining, The Proceedings of ACL'99: The 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland, June 20–26, 1999.

[16] R.N. Kostoff, H.J. Eberhart, D.R. Toothman, Database tomography for technical intelligence: a roadmap of the near-earth space science and technology literature, Inf. Process. Manag. 34 (1) (1998) 69–85.

[17] R.N. Kostoff, H.J. Eberhart, D.R. Toothman, Hypersonic and supersonic flow roadmaps using bibliometrics and database tomography, J. Am. Soc. Inf. Sci. 50 (5) (1999 (15 April)) 427–447.

[18] R.N. Kostoff, T. Braun, A. Schubert, D.R. Toothman, J.A. Humenik, Fullerene roadmaps using bibliometrics and database tomography, J. Chem. Inf. Comput. Sci. 40 (1) (2000 (Jan–Feb)) 19–39.

[19] R.N. Kostoff, K.A. Green, D.R. Toothman, J. Humenik, Database tomography applied to an aircraft science and technology investment strategy, J. Aircr. 37 (4) (2000 (July–August)) 727–730.

[20] R.N. Kostoff, R.A. DeMarco, Science and technology text mining, Anal. Chem. 73 (13) (2001 (1 July)) 370A–378A.

[21] R.N. Kostoff, R. Tshiteya, K.M. Pfeil, J.A. Humenik, Electrochemical power source roadmaps using bibliometrics and database tomography, J. Power Sources 110 (1) (2002) 163–176.

[22] R.N. Kostoff, M. Shlesinger, R. Tshiteya, Nonlinear dynamics roadmaps using bibliometrics and database tomography, Int. J. Bifurc. Chaos 14 (1) (2004 (January)) 61–92.

[23] R.N. Kostoff, M. Shlesinger, G. Malpohl, Fractals roadmaps using bibliometrics and database tomography, Fractals 12 (1) (2004 (March)) 1–16.

[24] R.N. Kostoff, G. Karpouzian, G. Malpohl, Text mining the global abrupt wing stall literature, J. Aircr. 42 (3) (2005) 661–664.

[25] R.N. Kostoff, R. Tshiteya, K.M. Pfeil, J.A. Humenik, G. Karypis, Power source roadmaps using database tomography and bibliometrics, Energy 30 (5) (2005) 709–730.

[26] R.N. Kostoff, H.J. Eberhart, D.R. Toothman, Database tomography for information retrieval, J. Inf. Sci. 23 (4) (1997).

[27] R.N. Kostoff, Systematic Acceleration of Radical Discovery and Innovation in Science and Technology, DTIC Technical Report, vol. ADA430720, Defense Technical Information Center, Fort Belvoir, VA, 2005, http://www.dtic.mil/.

[28] R.N. Kostoff, J.A. Block, Factor matrix text filtering and clustering, JASIST 56 (9) (2005 (July)) 946–968.

[29] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, R. Harshman, Indexing by latent semantic analysis, J. Am. Soc. Inf. Sci. 41 (6) (1990 (Sept)) 391–407.

[30] R.N. Kostoff, Research impact quantification, R&D Management 24 (3) (1994 (July)).

[31] Kostoff, R.N. Method for data and text mining and literature-based discovery. United States Patent Number 6,886,010. 26 April 2005.

[32] R.N. Kostoff, R.R. Schaller, Science and technology roadmaps, IEEE Transactions on Engineering Management 48 (2) (2001 (May)) 132–143.
[33] R.S. Burt, Structural Holes: The Social Structure of Competition, Harvard University Press, Cambridge, MA, 1992.

Ronald N. Kostoff received a PhD in Aerospace and Mechanical Sciences from Princeton University in 1967. At Bell Labs from 1966–1975, he performed technical studies in support of the NASA Office of Manned Space Flight, and economic and financial studies in support of AT&T Headquarters. At the U.S. Department of Energy from 1975–1983, he managed the Nuclear Applied Technology Development Division, the Fusion Systems Studies Program, and the Advanced Technology Program.

At the Office of Naval Research from 1983–present, he was Director of Technical Assessment for 10 years. He invented and patented (1995) the Database Tomography process, a computer-based textual data mining approach that extracts relational information from large text databases, and recently received a patent for a full-spectrum text mining system. After managing the Navy Laboratory Independent Research Program for five years, he established a new effort in textual data mining. His interests continue to revolve around improved methods to assess the impact of science and technology, incorporating maximal use of the massive amounts of data available.