Available online at www.sciencedirect.com
ScienceDirect
Technological Forecasting & Social Change 75 (2008) 186–202
Technological Forecasting and Social Change

Literature-related discovery (LRD): Methodology ☆

Ronald N. Kostoff [a,*], Michael B. Briggs [b], Jeffrey L. Solka [c], Robert L. Rushenberg [d]

[a] *Office of Naval Research, 875 N. Randolph St., Arlington, VA 22217, USA*
[b] *Arlington, VA 22204, USA*
[c] *Naval Surface Weapons Center Dahlgren Division, Dahlgren, VA 22448-5100, USA*
[d] *DDL-OMNI Engineering, LLC, 8260 Greensboro Drive, McLean, VA 22201, USA*

Received 30 July 2007; received in revised form 26 October 2007; accepted 31 October 2007

Abstract

Literature-related discovery (LRD) is linking two or more literature concepts that have heretofore not been linked (i.e., disjoint), in order to produce novel, interesting, plausible, and intelligible knowledge. LRD has two components: Literature-based discovery (LBD) generates potential discovery through literature analysis alone, whereas literature-assisted discovery (LAD) generates potential discovery through a combination of literature analysis and interactions among selected literature authors. In turn, there are two types of LBD and LAD: open discovery systems (ODS), where one starts with a problem and arrives at a solution, and closed discovery systems (CDS), where one starts with a problem and a solution, then determines the mechanism(s) that links them.

The generic methodology for identifying potential discovery candidates through ODS LRD, focusing mainly on its ODS LBD component, is described in this paper. A comprehensive flow chart showing the details of our systematic potential discovery generation process, including the evolution of the flow chart steps through each of the studies performed, is presented. Also shown is a vetting procedure that insures potential discoveries claimed are potential discoveries realized. The semantic filters that replace the numerical filters of other ODS LBD approaches are overviewed. The rationale for addressing the five topics studied (Raynaud's Phenomenon (RP), Cataracts, Parkinson's Disease (PD), Multiple Sclerosis (MS), and Water Purification (WP)) is summarized.

Keywords: Literature-Based Discovery; Text Mining; Information Retrieval; Clustering; Semantic Filters

☆ DISCLAIMER: (The views presented in this paper are those of the authors, and do not represent the views of the Department of the Navy or any of its components, or DDL-OMNI, LLC).

* Corresponding author.
E-mail address: kostofr@onr.navy.mil (R.N. Kostoff).

0040-1625/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.techfore.2007.11.010

1. A systematic approach for accelerating discovery and innovation

1.1. Process requirements for discovery

The purpose of this paper is to identify characteristics of potential discovery, and to present a generic approach for targeting potential discovery. Equally important, as demonstrated by some of the problems in the Background section of the introductory paper [1], a vetting approach is described that insures verification of claimed potential discovery.

All LBD/LAD types share the common feature that more than one literature is required to address the problem of interest. If only one literature were required, then the solution would have been discovered by the producers and readers of that literature. What properties should these multiple literatures have for credible LBD/LAD?

- All the literatures contributing toward the solution of the problem should be complementary. Each literature should contain unique information that contributes to the total problem solution, and without each literature's unique contribution the overall problem cannot be solved.
- All the literatures should be disjoint. Otherwise, an individual's or group's knowledge of all literatures would eliminate literature-based discovery, since the information contributing to the solution could be pieced together by one individual.
- All these literatures should be as comprehensive as possible; otherwise, the disjoint-ness assumption may be a consequence of the limited literature selected, and may not be valid.
- All these literatures with unique information must be linked to form a whole that is greater than the sum of its parts.

The first author's text mining efforts over the past decade have been focused on 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. Our group has applied text mining to assessing the technical structure and infrastructure of 1) single technologies [e.g., nanotechnology, anthrax] [2,3] and 2) country portfolios of myriad technologies [China–India] [4–7]. These methods have been integrated with the discovery literature characteristics above to form a systematic approach for accelerating discovery.

In particular, we have developed a generic approach to systematic acceleration of ODS LRD, and have applied six variants of this approach (mainly ODS LBD variants) to five problems: four medical (RP, cataracts, PD, MS) and one physical science (WP). After summarizing the generic approach, we will proceed in succeeding papers to the details of the approach and the potential discoveries made on the five problems.

1.2. Summary of generic approach to ODS LRD

1. Retrieve core literature of target problem
   - Generate query for core literature
   - Enter query into database search engine and retrieve core literature
2. Characterize core literature
   - Obtain technical infrastructure of core literature (key researchers, Centers of Excellence) through bibliometrics