Available online at www.sciencedirect.com
ScienceDirect
Technological Forecasting & Social Change 75 (2008) 256–275
Technological Forecasting and Social Change

Literature-related discovery (LRD): Water purification

Ronald N. Kostoff [a,*], Jeffrey L. Solka [b], Robert L. Rushenberg [c], Jeffrey A. Wyatt [c]

[a] *Office of Naval Research, 875 N. Randolph St., Arlington, VA 22304, USA*
[b] *Naval Surface Weapons Center Dahlgren Division, Dahlgren, VA 22448-5100, USA*
[c] *DDL-OMNI Engineering, LLC, 8260 Greensboro Drive, McLean, VA 22102, USA*

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

Abstract

Literature-related discovery (LRD) is the linking of two or more literature concepts that have heretofore not been linked (i.e., disjoint), in order to produce novel, interesting, plausible, and intelligible knowledge (i.e., potential discovery). LRD has two main components that differ in their methodological approach to discovery: Literature-based discovery (LBD) produces potential discovery through analysis of the technical literature alone; Literature-assisted discovery (LAD) produces potential discovery through both analysis of the technical literature and use of selected authors of that literature. These authors generate potential discovery as proposers, workshop/panel participants, or in other active roles.

The open discovery systems (ODS) component of LRD starts with a problem to be solved, and generates solutions to that problem through potential discovery. We have been using ODS LRD to identify potential treatments or preventative actions for challenging medical problems, among myriad other applications. The previous four papers in this Special Issue describe the application of ODS LRD (specifically, the ODS LBD variant) to Raynaud's Phenomenon (RP), cataracts, Parkinson's Disease (PD), and Multiple Sclerosis (MS).

One goal of the present study was to determine whether LRD could be successfully applied (for the first time) to a challenging non-medical technical problem to generate potential discovery. The second goal was to explore the use of both LRD variants (LBD and LAD) to a non-medical technical problem. We selected the problem of water purification (WP) because of universal applicability and sponsor interest.

We used LRD to identify purification concepts, technology components and systems that could lead to improved water purification techniques. We accessed many disparate disciplines to identify purification concepts from literatures not normally associated with water purification. We used two LBD approaches, Cluster Filtering and Latent Semantic Indexing (LSI), to search for potential discovery. We generated voluminous amounts of potential discovery, and believe we have only scratched the surface of what is possible. We also ran a short

* Corresponding author. Fax: +1 703 696 8744.
E-mail address: kostofr@onr.navy.mil (R.N. Kostoff).

doi:10.1016/j.techfore.2007.11.009

experiment using LAD to identify experts associated with potential discovery concepts, and use their expertise to generate potential discovery for water purification.

Keywords: Literature-related discovery; Literature-based discovery; Literature-assisted discovery; Text mining; Cluster filtering; Latent semantic indexing; Water purification; Drinking water

1. Overview of study

The previous four papers in this Special Issue describe the application of ODS LRD (specifically, the ODS LBD variant) to Raynaud's Phenomenon (RP) [1], cataracts [2], Parkinson's Disease (PD) [3], and Multiple Sclerosis (MS) [4]. The extremely positive results from the RP study (potential treatments) spawned a larger study on water purification (WP). We selected the specific problem of WP because of universal applicability and sponsor interest. Its objectives were to identify potential improvements in the water purification process, especially (but not limited to) cost reduction. Two LBD approaches were selected: Cluster Filtering and Latent Semantic Indexing (LSI). One LAD approach was selected: Broad Agency Announcement Notification.

Cluster Filtering was inspired by the use of cluster analysis in query development. When we clustered retrievals of records relevant and non-relevant to some topical query, we noticed that the clustering tended to segregate the relevant from the non-relevant records. As stated in the first author's patent: "Document clustering tends to group documents into groups that are at similar levels of relevance. A technical expert then samples documents from each group, and performs a final judgment as to the relevance of each group." [5]. The reason for this segregation was that the relevant records would contain a number of terms related to the query topic, whereas the non-relevant records would typically contain one query term. In the article's context, the term tended to be peripheral to the theme of the query. The non-relevant records would form into clusters each of which centered about the one query term, and the relevant records would multi-link cluster around the multiple terms. We hypothesized that this relevant/non-relevant segregation could be extrapolated to searching for discovery. However, the remainder of the Cluster Filtering approach was conceptually similar to that used in the RP study, and was manually intensive.

LSI-based discovery was an attempt to use more sophisticated algorithms to help identify potential discovery candidates. It allowed much easier access to very remotely related literatures compared to the manually intensive methods described above.

Even though the Cluster Filtering approach started chronologically much later than the LSI approach, the Cluster Filtering approach will be described first. It is more closely related to the RP approach than is the LSI approach. The LAD approach will be summarized briefly at the end.

2. LBD approaches and results

2.1. Cluster Filtering

2.1.1. Background of water purification

Water purification is the process of removing contaminants from a raw water source. The goal is to produce water for a specific purpose with a treatment profile designed to limit the inclusion of specific