**Preface**

This book is intended to address the earlier part of the process in developing a research strategy. The challenge is to transform the personal, private deliberations leading to critical and creative thinking to a more transparent process. If that can be made more public by using computer algorithms, the innovative paths can be better understood. The concepts involved are comparable to the proof of theorems. The innovator develops a theorem and its proof. Students learn to duplicate the process leading to the desired result and are introduced to the process of critical and creative thinking in that way.

Typically, the research process is presented beginning with the recognition of a problem to be addressed. The process leading to that recognition is hidden. There are different ways to select a problem. It could be recognized after a review of the existing literature. It could be determined by observation during daily activities. It could be determined after discussion with colleagues. The particular path leading to the recognition may be hidden from the individual so that when queried, the best response is – the problem is important ***because…***

Making the critical and creative thinking process transparent begins with an important mechanical task – the identification, extraction, and organization of the authors’ ideas from scholarly publications. That task can be performed by software. The result is a separation of the mechanical and intellectual components of building and using a resource. The essential data consist of authors’ ideas. These intellectual building blocks can be used to develop descriptions of the topic and to identify gaps or inconsistencies that could lead to new research strategies.

The use of ideas can be made transparent by application of computer algorithms. The advantage is a discernible path of tasks describing the critical or creative act. This separation of mechanical acts performed using algorithms and intellectual acts defining order of and intent of tasks has advantages. Among which is a shortened time and a shift in energy expenditure from the mechanical/clerical to the intellectual. The emphasis is on the higher cognitive functions – synthesis, comparison, evaluation, judgment, and application. The focus is on the development of ***measures*** associated with each cognitive function, the establishment of ***criteria*** to determine how to use those measures, and the recognition of ***decision-rules*** describing the resulting actions associated with the measures and criteria. These deliberations are examples of critical thinking and intellectual function.

The application of text mining software to scholarly publications is not without challenges. As an example, the study of dog-related diseases involves contributions from epidemiology, clinical, and laboratory science. As such, the text mining software must deal with different writing styles and vocabularies. In addition to the variation in specific topics considered and ways of describing them, there are numerous sources contributing bibliographic data to PubMed. This bibliographic resource contains over 24 million documents making the assessment of accuracy a monumental task. The resource is continuously changing making retrieval results an estimate rather than a fact. Given these conditions, the search engine identified 27,926 documents (1980-2013) or 79% of the documents containing the phrase – dog disease – in their list of subject headings. This capture resulted in almost 3 million ideas.

A significant advantage of the text mining approach is the ability to provide estimates of completeness and accuracy. The matching of vocabulary identified (nouns, adjectives, and gerunds) used by authors with that identified by the software yielded a median of 85% (66% to 99%) across subjects. In terms of capture of author ideas, the software identified over 90% of those used. These estimates of vocabulary identification compared favorably with other approaches – human indexing identified 50% - 60% of the authors’ vocabulary. Other text mining methods identified about 30% and random sampling yielded about 20%. Idea capture was not determined by methods other than the Idea Analysis system.

**Table 1. Occurrence of Cancer Related Ideas – 1980-2013.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Term** | **1980-4** | **1985-9** | **1990-4** | **1995-9** | **2000-4** | **2005-9** | **2010-13** |
| breast | 8 | 6 | 18 | 7 | 13 | 51 | 33 |
| prostate | 0 | 0 | 1 | 20 | 19 | 34 | 19 |

Table 1shows the occurrence of cancer and breast or prostate ideas. The peak use of the ideas was in interval 2005-9. Prostate wasn’t linked with cancer until the period 1990-1994.

**Figure 1. Occurrence of Cancer with Personal Factors – 1980-2013.**

Figure 1 shows the frequency of use of ideas linking cancer with different personal dimension terms. Of those, use of the ideas – cancer & dog or cancer & canine – increased to a high in the interval 2005-9. The idea – cancer & animal – showed a minimal increase in the same period. The idea – cancer & breed – was used infrequently.

In addition to the eBook describing the earlier part of research development, a comprehensive idea database providing ideas from 1980 – 2013 is included. These ideas are organized as excel files and may be downloaded for detailed processing. In searching for new problems to study or the development of new descriptions of existing knowledge, it is useful to begin with the most recent ideas (2010-2013). When specific ones are identified, their occurrence in earlier years can be determined and a temporal description of the ideas developed. This approach is useful in identifying new ideas versus ones that have been considered for longer periods of time.

The advantage of an idea database using Excel is that the use of the software is widely known. The primary functions are:

1. Selection of record subsets.
2. Sorting of records.
3. Copying of records.
4. Counting of records.

By combining these functions, existing and new arrangements of ideas can be rapidly developed. As a result, the emphasis is on measures, criteria, and decision-rules associated with comparing the various syntheses.

John M. Weiner, Dr.P.H.

[weiner.john@tutorghost.com](mailto:weiner.john@tutorghost.com)

William McAfoos,BCE

[wmcafoos@btbuffalo.com](mailto:wmcafoos@btbuffalo.com)