**Chapter 2 -- New Tools in Investigation and Learning**

**Introduction**

***Literature Analysis*** differs from the well-known ***Literature Review*** by requiring that each of the tasks involved in identifying, extracting, organizing, and disseminating information is transparent and subject to quality control. Literature Review involves personal, private, mentation. The time required to learn this process is long and the quality is uncertain. The only way to verify is to repeat the private process. Discrepancies in results can be attributed to fatigue and the use of shortcuts rather than changing data. If the data are stable, the processing and results should be the same each time. That can be accomplished only by employing public, quality-monitored tasks, the intent of Literature Analysis.

**Dealing with the Actual Volume of Ideas:** There is a significant drawback associated with identifying, extracting, and organizing the authors’ ideas contained in documents. That is the larger volume that must be considered. There also are significant advantages associated with the Literature Analysis approach. One is the ability to employ quality control procedures in the construction of the idea database as well as in the accomplishment of the higher cognitive functions. That is, the intellectual tasks required can be made more transparent, thus enhancing the actual procedures used by the specialist. This oversight can be beneficial to the student who is challenged to accomplish the transformation to professional. Each function can be subdivided into three key areas, namely, the development of – ***measures/observations*** describing the phenomenon considered, ***criteria*** using those variables, and ***decision-rules*** leading to resulting actions.

**The Question and Resulting Idea Sets**: The question of interest is – ***Can formal identification and organization of the ideas presented by specialists in their scientific documents be used to effectively assess the current knowledge and to recognize omissions?*** This version of gap analysis (a method to identify omissions in the body of knowledge) could be useful in the description of existing knowledge and the development of new knowledge.

Using the ideas as an evaluative basis, four sets of ideas could be recognized. These are:

1. ***New ideas*** that enhance understanding of the phenomena representing the subject. These ideas would occur with low frequency in the total set and later in time. That is, in temporal analyses, these ideas would appear to be emerging.
2. Ideas presenting ***confirmatory evidence of concepts*** in a new environment. These ideas occur with higher frequency and throughout the period of study. The environment would be a new one and the relevance of the findings is in the replication across environments.
3. Ideas revealed by ***new methods*** providing improved insights. New methods can open doors to new opportunities and measures. The ideas may be well established but the measures involved in depicting them can be new.
4. ***Existing ideas*** in previously studied environments. These ideas represent a type of consensus of accepted knowledge. This set represents the background of existing knowledge.

This classification of ideas, based on frequency and time of occurrence, facilitates development of existing knowledge and strategies for future knowledge acquisition. These outcomes have a long tradition of intellectual accomplishment. The methodological distinctions offered by Literature Analysis and Gap Analysis are transparency, accuracy, and efficiency. Using those methods, innovators can accomplish their objectives faster. The more public process enables students to learn and duplicate these innovations.

**Methods**

The Literature Analysis method is based on the premise that software can accomplish the following tasks.

1. Identify and separate the authors’ sentences so that each can be considered as a domain containing informative terms (nouns, adjectives, or gerunds) combined by the authors to present thoughts or ideas.
2. Identify each pair of informative terms (i.e., idea) within each sentence.
3. Prepare a data record consisting of the idea and the bibliographic data describing the document and its location.
4. Organize those data records as a file for subsequent use.

The document set containing the ideas dealing with dental care in chronic disease numbered 18,444 and yielded 1,231,780 ideas based on combinations of 1554 informative terms. The software identified informative terms using defining suffixes and contextual meaning of terms. The time required to identify, extract, and organize the ideas was 0.3 minutes per 250 word document and involved three readings of each document. The first reading identified informative terms. The second prepared the idea records. The third confirmed these records.

The software identified a median of 85% of the informative terms (66% to 99%) employed by the author-specialists. Over 95% of the ideas were correctly identified and included in the idea database. The identification of informative terms compared favorably with previous approaches – human indexing (50% - 70%), statistical text mining methods (30% - 50%), and random selection (10% - 20%). The processing time was significantly less than the previous methods including text mining designed to cluster documents into themes. That approach required manual inspection of the clusters to validate document membership.

Exhibit 1 shows the result of the identification, extraction, and organization process. The informative terms within the first sentence are highlighted in red. These terms are arranged as pairs and entered into a data file. The idea records show these pairs (ideas) together with the location of the sentence. The identification number was assigned by PubMed and provides a portal for rapid retrieval of the document of interest. This identification and extraction process is effective in capturing the ideas expressed by the authors and facilitates use of those in building descriptions of existing knowledge (higher frequency and consistently expressed ideas) as well as strategies for development of new research (the gap analysis using lower frequency, current ideas).

**Exhibit 1. Example of Sentence Containing Ideas and the Ensuing Data Records Included in the Idea Repository – Identification Number 19151554.**

**Source:** [**Silva AR**](http://www.ncbi.nlm.nih.gov/pubmed?term=Silva%20AR%5BAuthor%5D&cauthor=true&cauthor_uid=19151554)**1,** [**Alves FA**](http://www.ncbi.nlm.nih.gov/pubmed?term=Alves%20FA%5BAuthor%5D&cauthor=true&cauthor_uid=19151554)**,** [**Antunes A**](http://www.ncbi.nlm.nih.gov/pubmed?term=Antunes%20A%5BAuthor%5D&cauthor=true&cauthor_uid=19151554)**,** [**Goes MF**](http://www.ncbi.nlm.nih.gov/pubmed?term=Goes%20MF%5BAuthor%5D&cauthor=true&cauthor_uid=19151554)**,** [**Lopes MA**](http://www.ncbi.nlm.nih.gov/pubmed?term=Lopes%20MA%5BAuthor%5D&cauthor=true&cauthor_uid=19151554)**. Patterns of demineralization and dentin reactions in radiation-related caries.** [**Caries Res.**](http://www.ncbi.nlm.nih.gov/pubmed/?term=19151554) **2009;43(1):43-9. doi: 10.1159/000192799. Epub 2009 Jan 19. PMID: 19151554.**

**Sentence 1:** ***Radiation-related caries is a unique form of rampant decay and is a complication of head and neck radiotherapy that frequently causes generalized dental destruction and impairs quality of life in cancer patients.***

**Idea Records:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Primary** | **Related** | **Year** | **Ident** | **Sentence** |
| **cancer** | **caries** | **2009** | **19151554** | **1** |
| **cancer** | **cause** | **2009** | **19151554** | **1** |
| **cancer** | **dental** | **2009** | **19151554** | **1** |
| **cancer** | **life** | **2009** | **19151554** | **1** |
| **cancer** | **radiation** | **2009** | **19151554** | **1** |
| **cancer** | **radiotherapy** | **2009** | **19151554** | **1** |

**Results**

There were 1,231,953 ideas retrieved from PubMed for the period 1980-2013. The total number of ideas in the decade 1980-89 was in excess of 74,000. The four year period – 2010-2013 – yielded over one-half million ideas. Age was the term involved in the largest number of ideas, 47,211. This was followed by cell (47,139), oral (46,750), disease (41,390), dental (36,616), and gene (30,938). The first disease identified was cancer with 28,318 ideas. Periodontal and periodontitis were the first dental conditions identified with respectively, 20,900 and 13,286 ideas.

**Caries Management:** The next dental condition identified was caries with 6,544 ideas. This disease presents a problem in health management and the ideas involved will be used to illustrate the procedures involved in using the new tools – Literature Analysis and Gap Analysis.

**Figure 1. Higher Frequency Terms Linked with the Central Term – Caries – 1980-2013.**

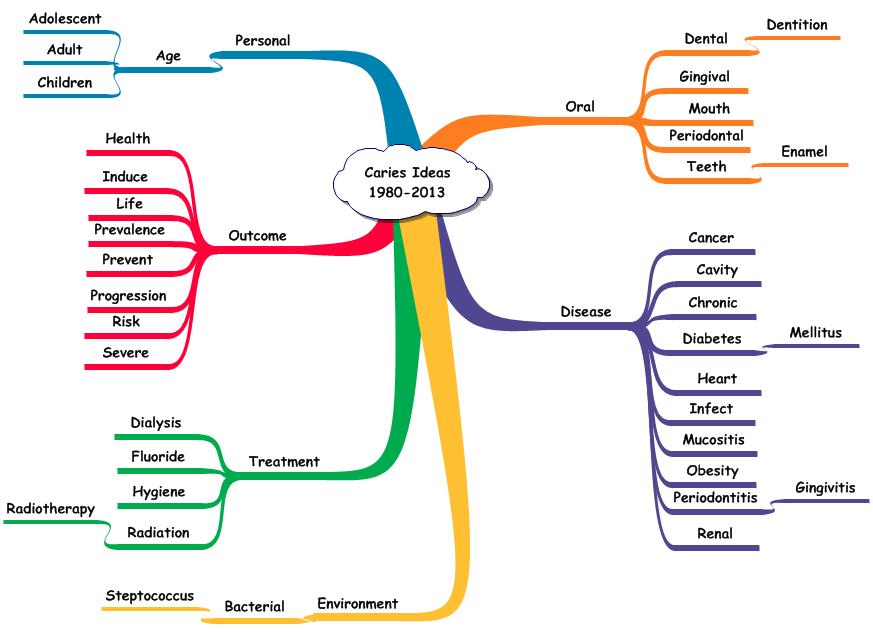
****

Figure 1 shows the higher frequency terms linked with the central term – ***caries*** – for the period 1980 – 2013. There are several clusters of ideas. One is the ideas describing oral issues associated with caries. Those ideas focus on dental structural aspects. Another cluster described the diseases associated with caries. These could be summarized as including immuno-compromised patients. A third cluster described the results of management and included ideas dealing with status, failure and, of particular interest, prevention. A fourth cluster dealt with treatments. Radiation therapy and dialysis were relevant in occurrence of caries in patients with oral cancer or kidney disease. The other treatments identified – hygiene and fluoride -- dealt with approaches in attempting to deal with caries occurrence.

**Ideas Involving Prevent:** The idea – ***Caries and Prevent*** – is of interest in determining what treatments have been explored in minimizing or preventing oral infections in the light of chronic diseases. Figure 2 shows the terms linked with the central outcome term – ***prevent*** – during the period 1980-2013. Current (2010-13) diseases linked with the central term – prevent – were diabetes, heart, infection, and oral sites. Treatments included antibiotics, hygiene, prosthetics, and transplant.

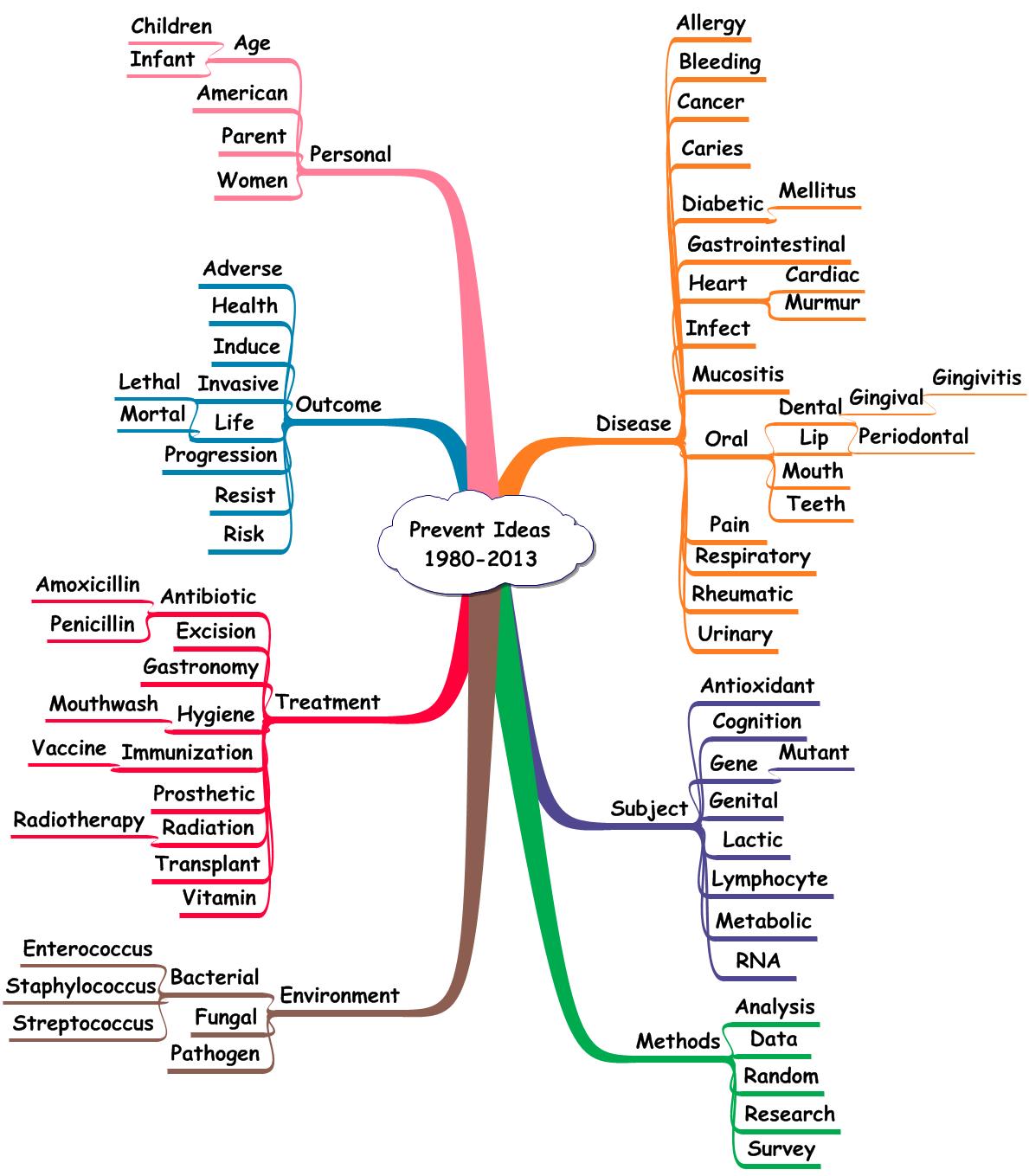
**Figure 2. Terms Linked with the Central Term – Prevent – 1980 -2013.**

Table 2 shows the dental treatment terms and the number of ideas involving each through the period 1980 through 2013. The ideas involving – antibiotics – were observed 204 times while those involving fluoride were present 84 times. Mouthwash (20), chlorhexidine (18), toothpaste (8), mouthrinse (4), and xylitol (4) were lower frequency related ideas.

**Table 2. Terms Linked with the Central Terms Depicting Dental Treatments.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Treatment** | **1980-9** | **1990-9** | **2000-4** | **2005-9** | **2010-13** | **Sum** |
| antibiotic | 16 | 17 | 51 | 18 | 102 | 204 |
| fluoride | 7 | 7 | 13 | 15 | 42 | 84 |
| implant | 3 | 1 | 11 | 20 | 35 | 70 |
| prosthetic | 6 | 2 | 7 | 4 | 19 | 38 |
| fluoridation | 2 | 5 | 1 | 4 | 12 | 24 |
| immunization | 1 | 1 | 5 | 4 | 11 | 22 |
| vaccination | 0 | 3 | 3 | 5 | 11 | 22 |
| vaccine | 3 | 1 | 4 | 3 | 11 | 22 |
| denture | 1 | 2 | 4 | 3 | 10 | 20 |
| endodontic | 0 | 6 | 1 | 3 | 10 | 20 |
| mouthwash | 0 | 3 | 1 | 6 | 10 | 20 |
| chlorhexidine | 1 | 1 | 1 | 6 | 9 | 18 |
| penicillin | 4 | 2 | 0 | 0 | 6 | 12 |
| catechin | 0 | 2 | 1 | 2 | 5 | 10 |
| prosthesis | 0 | 0 | 3 | 2 | 5 | 10 |
| excision | 0 | 1 | 2 | 1 | 4 | 8 |
| toothpaste | 0 | 1 | 2 | 1 | 4 | 8 |
| mouthrinse | 0 | 1 | 0 | 1 | 2 | 4 |
| xylitol | 2 | 0 | 0 | 0 | 2 | 4 |
| prosthodontic | 0 | 0 | 0 | 1 | 1 | 2 |

**Idea Analysis vs. Hyperlinking**: The combination of terms representing an intervention and a result describes the effect of clinical intent. Two ideas set the stage for assessment of the intervention. For example, the idea – ***prevent and caries*** – describes the disease condition and its change. The idea – ***prevent and antibiotic*** – describes the treatment and the effect on the disease condition. ***Prevent and*** ***caries*** – occurred 157 times and ***prevent and antibiotic*** occurred 102 times. However, the idea that is most informative would be the one that combines the three terms– ***prevent, caries, and antibiotic.*** That triad occurred once. Doron et al (PMID: 16597207) came close to combining the terms in two adjacent sentences dealing with the role of probiotics –

***“The strongest evidence for the clinical effectiveness of probiotics has been in their use for the prevention of symptoms of lactose intolerance, treatment of acute diarrhea, attenuation of antibiotic-associated gastrointestinal side effects and the prevention and treatment of allergy manifestations.***

***More research needs to be carried out to clarify conflicting findings on the use of probiotics for prevention of travelers' diarrhea, infections in children in daycare and dental caries, and elimination of nasal colonization with potentially pathogenic bacteria.”***

A reasonable supposition would be that the authors did not intend to combine the three terms as an idea of relevance to them. Instead, they appear to be listing a series of conditions and, in so doing, incidentally included the pairs of interest. These sentences also illustrate the problems with hyperlinking as a mechanism in combining terms. The fact that the terms of interest were placed in adjacent sentences did not imply that the authors considered the combination as a relevant thought.

**Exhibit 2. Example of Study Designed to Control Caries in Cancer Patients – PMID 21180287.**

**Source:** [**Patel M**](http://www.ncbi.nlm.nih.gov/pubmed?term=Patel%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21180287)**1,** [**Ndlovu NN**](http://www.ncbi.nlm.nih.gov/pubmed?term=Ndlovu%20NN%5BAuthor%5D&cauthor=true&cauthor_uid=21180287)**,** [**Owen CP**](http://www.ncbi.nlm.nih.gov/pubmed?term=Owen%20CP%5BAuthor%5D&cauthor=true&cauthor_uid=21180287)**,** [**Veale R**](http://www.ncbi.nlm.nih.gov/pubmed?term=Veale%20R%5BAuthor%5D&cauthor=true&cauthor_uid=21180287)**. Properties of a new mouthrinse for patients receiving radiation therapy.** [**SADJ.**](http://www.ncbi.nlm.nih.gov/pubmed/?term=21180287) **2010 Oct;65(9):410, 412-4. PMID: 21180287**

**Sentence 28: *The experimental mouthrinse was cost-effective and proved to have an antimicrobial effect and could be used safely to alleviate oral infections, desensitize teeth, improve oral hygiene and control dental caries in cancer patients after radiation therapy.***

**Ideas:**

# 

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| caries | cancer | 2010 | 21180287 | 28 |
| caries | dental | 2010 | 21180287 | 28 |
| caries | hygiene | 2010 | 21180287 | 28 |
| caries | infect | 2010 | 21180287 | 28 |
| caries | mouth | 2010 | 21180287 | 28 |
| caries | mouthrinse | 2010 | 21180287 | 28 |
| caries | oral | 2010 | 21180287 | 28 |
| caries | radiation | 2010 | 21180287 | 28 |
| caries | teeth | 2010 | 21180287 | 28 |

**Hygiene Ideas**: The term, ***prevent***, representing an outcome, may not be as informative as a term representing a treatment. One such term used by the authors was ***hygiene***. Exhibit 2 gives an example of a report dealing with the study of a new mouthrinse in controlling dental caries after radiotherapy in cancer patients.

The informative terms in the involved sentence are highlighted in red and the idea records from the database are shown. The idea – ***caries and hygiene*** – occurred seven times. Of those reports, the one in Exhibit 2 was unique in studying a potential intervention program. As such, this low frequency representation of a treatment may point to a gap in the knowledge structure. Namely, ***in dealing with patients with chronic diseases (cancer, diabetes, heart, or kidney), what is the optimum treatment approach in minimizing the adverse effects of caries?***

**Table 3. Treatment Terms Linked with Caries – 1980-2013.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Term** | **1980-9** | **1990-9** | **2000-4** | **2005-9** | **2010-13** | **Total** |
| fluoride | 17 | 10 | 23 | 21 | 28 | 99 |
| hygiene | 7 | 5 | 12 | 17 | 38 | 79 |
| fluoridation | 4 | 2 | 5 | 1 | 2 | 14 |
| vaccine | 5 | 0 | 3 | 2 | 1 | 11 |
| denture | 0 | 1 | 2 | 4 | 3 | 10 |
| endodontic | 0 | 2 | 1 | 2 | 4 | 9 |
| prosthetic | 0 | 0 | 3 | 2 | 2 | 7 |
| amalgam | 0 | 0 | 0 | 6 | 0 | 6 |
| implant | 0 | 0 | 3 | 0 | 3 | 6 |
| antibiotic | 0 | 0 | 2 | 2 | 1 | 5 |
| toothpaste | 1 | 0 | 1 | 1 | 2 | 5 |
| chlorhexidine | 0 | 2 | 1 | 1 | 0 | 4 |
| mouthrinse | 1 | 0 | 1 | 0 | 1 | 3 |
| xylitol | 2 | 0 | 1 | 0 | 0 | 3 |
| prosthesis | 0 | 0 | 0 | 1 | 1 | 2 |
| prosthodontic | 0 | 1 | 0 | 1 | 0 | 2 |
| catechin | 0 | 0 | 1 | 0 | 0 | 1 |
| mouthwash | 0 | 0 | 1 | 0 | 0 | 1 |
| toothbrushing | 0 | 0 | 0 | 1 | 0 | 1 |

The treatments studied with – ***caries*** – are shown in Table 3. The number of ideas involving caries ranged from – caries and fluoride (99) to caries and toothbrushing (1). The low frequency treatments studied in the current period (2010-2013) included prosthesis, mouthrinse, toothpaste, and antibiotic. Each could be a possible inclusion in the optimal treatment program.

**Discussion**

This report introduces an alternative approach involving:

1. Introduction of transparency in performing each text processing task.
2. Employment of quality control of the process so that accuracy and completeness can be determined.
3. Development of a knowledge resource that can serve all investigators.
4. Facilitation of rapid, accurate, and comprehensive use of the knowledge.

The shift in processing by assigning each task to the most efficient and effective capability implies that the time-consuming search, retrieval, and organization of data can be addressed once. With that resource verified and validated, the time and energy expended by numerous investigators can be shifted from the clerical/mechanical functions to the intellectual ones – synthesis, comparison, evaluation, judgment, and application. In addition, these heretofore private and personal tasks can be made more public with quality control monitoring for accuracy and completeness.

**Array of Syntheses:** The array of possibilities is not unfamiliar. Automobile manufacturers display their various models and invite assessment of numerous measures of performance, size, and quality. This assessment is intended to lead to the selection of the one auto satisfying the criteria held important by the user. Similarly, household products, wearing apparel, and numerous other products are offered in this comparative fashion. The one important product that is not presented in this way is the array of organized ideas. Presumably, manual processing is too restrictive to consider performance of the tasks more than once. The focus on ideas and the use of computer algorithms changes that capability to one of formalized learning and knowledge utilization using efficient methods.

**Stages of Learning**: The Idea Analysis approach facilitates developing a description of the phenomenon of interest by relying on the importance assigned by subject specialists to relationships (Stage 1). With that comprehensive view of the idea structure, gaps in knowledge are easier to identify (Stage 2). The investigator continues to have the important decision-making with respect to which new arrangement of ideas will yield the greatest amount of definitive new knowledge (Stage 3). To assist him in that challenge, the algorithms that enable comparison and evaluation speed the time to conclusion. Algorithms that rank the selected syntheses on the basis of selected measures facilitate the investigator’s choice of the ‘best’ one (Stage 4). Translating that idea structure into a data-gathering process is well established, as is the analysis of the findings. Those results should lead to a revised idea structure containing the previous and the new (Stage 5).

**Summary**

Newer forms of learning (problem-based, on-demand, discovery) require a change in approach and the processing tools used. An alternative set of tools was considered in dealing with such situations. The process was illustrated using the documents describing dental issues in chronic diseases. There were over 1 million ideas in that topic and of those over 6000 described caries related ideas. The syntheses developed included exploration of caries related ideas involving the chronic diseases studied, the treatments employed, the changes associated with study observations, and methods used.

In addition to describing the findings experienced by subject specialists, the analytic approach can identify gaps in the idea structure. Depending on the investigator’s assessment of importance, specific gaps could be the basis for new knowledge generating studies. This process of translating omissions in the idea structure to formalized research studies is a form of Gap Analysis.

**Advantages:** The combination of reliance on the world’s experts for ideas (Literature Analysis) and identification of omissions in previous findings (Gap Analysis) has the added virtues of speed in accomplishment, determination of accuracy and completeness, and transparency so that the intellectual process can be replicated by others.