**Chapter 3 -- Idea Analysis – A Research Tool**

**Introduction**

Given the exponentially increasing amount of information available, the individual-- student or professional -- is faced with the challenge of identifying, extracting, organizing, and utilizing facts and opinions to form new descriptions of a topic. Research plays an important role in dealing with the deluge of information. In a real sense, research is a formalized learning process, pushing back the unknown and discovering aspects of new topics.

Many books have been written describing research methods. Few however have focused on the full range of tasks and how they can be performed using the most effective capabilities. Those tasks are:

1. ***Developing a learning resource***. This repository of essential data provides the ingredients needed to construct descriptions of the topics and to develop new strategies leading to advancing knowledge. This construction process can best be accomplished by software designed to identify, extract, and organize specific data from the scientific publications describing the topic. By using software, there is enhanced transparency of procedure. The measures, criteria, and decision-rules must be specified in advance and use of those performed in a consistent fashion.
2. ***Using the learning repository.*** The data in the learning resource represent the totality of conceptual information regarding the topic. The student can rapidly learn the basic description of the topic by focusing on the information considered most relevant by the subject specialists who developed it. This specialist-guided mentoring and instruction expands on the work of an author of a textbook by providing a worldwide view of the topic rather than an individual’s interpretation.

The professional can rapidly access the subsets of data considered relevant in addressing a specific topic. In addition to saving time and effort, the subsets selected will be more complete than an individual’s personal perspective of the knowledge, thus providing more opportunities to expand the body of relevant information.

1. ***Formalized learning.*** With the learning repository available, the individual’s time and energy can be shifted from ***acquiring*** data to ***using*** the higher cognitive functions. The first of these is synthesis, the construction of new descriptions of the topic or issues in the topic based on arrangements of the essential data from the learning repository. In contrast to previous methods, this version of synthesis would lead to many different possible arrangements. This array of new descriptions offers the opportunity to compare, evaluate, and judge each leading to a ranking denoting most potentially informative to least.

The selected application could be called a testable hypothesis or the question to be studied in a research program. It is the ***best*** arrangement of the existing facts and represents the most plausible portal to new understandings. Those results will be determined by the new study. In this sense, the process of developing a study hypothesis and the process involved in populating that hypothesis with new information are transparent, quality-controlled events. Learning accomplished using this approach can be described as a system of formalized performance of specific tasks, operationalized critical thinking, or simply as research.

**Separating the Processing Tasks:**  The approach described in this discussion, i.e., idea analysis, using computer support, allows separation of the individual tasks and assigns each to the most capable and effective methodology. The implication of this separation is the ability to consider the details and issues associated with each cognitive function. This emphasis leads to a more open process and could lead to a formalized version of critical thinking and its companion, creative thinking.

Critical and creative thinking are recognized desirable skills applied in various ways in higher education. However, a significant challenge is the translation of this behavior from an ideal to an operational entity.

**Example from Dental Cancer Research:** Table 1 shows the different terms describing cancer studied and the frequency of ideas involving each term for the period 2010-2013.

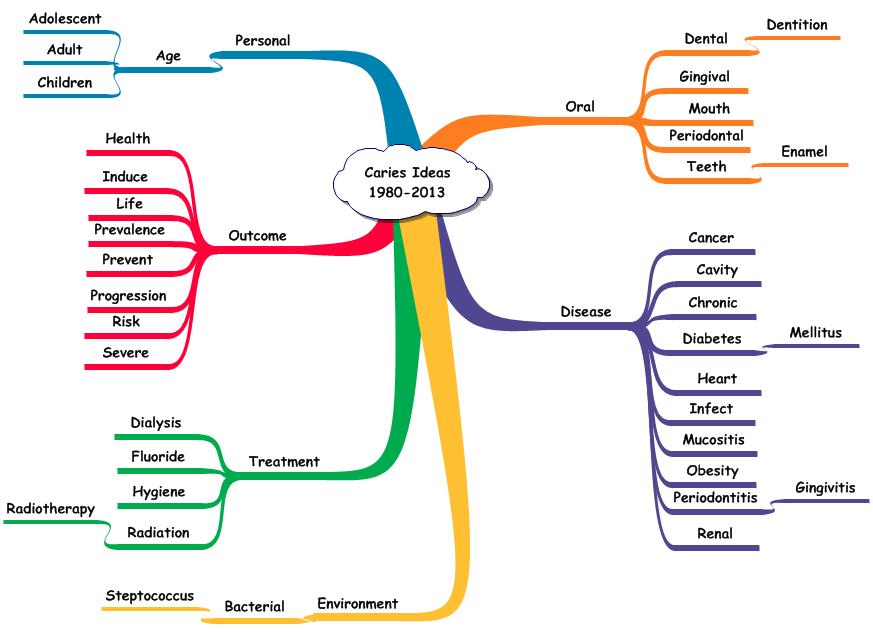
**Table 1. Excerpt of Cancer Ideas From Dental Disease Vocabulary – Cancer Ideas 2010-13.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Term** | **Freq** |  | **Term** | **Freq** |
| **2010-13 Cancer Ideas** | **88439** |  | **dental** | **572** |
| **tumor** | **10670** |  | **melanoma** | **543** |
| **malignant** | **3832** |  | lesion | 524 |
| oral | 2498 |  | **benign** | **496** |
| **metastasis** | **2483** |  | **breast** | **478** |
| cell | 2369 |  | **carcinogen** | **467** |
| **malignancy** | **2147** |  | clinical | 448 |
| **neoplasm** | **1501** |  | bone | 437 |
| age | 1300 |  | **malignancies** | **429** |
| gene | 1144 |  | life | 406 |
| **neoplastic** | **783** |  | tissue | 357 |
| disease | 748 |  | squamous | 355 |
| treatment | 718 |  | cause | 351 |
| risk | 649 |  | health | 349 |
| therapy | 626 |  | progression | 339 |
| **metastases** | **608** |  | stage | 339 |
| **sarcoma** | **606** |  | resection | 329 |

The large number of ideas (e.g., 88,439 cancer ideas) that must be processed to find the ones that are relevant is a condition of text processing. The only way to accurately describe each document is by identifying the authors’ ideas. To do that means that each document must be read, the ideas extracted and organized for future use. The use of software enabled this effective triaging in a rapid and efficient fashion.

**Idea Maps:** Graphics showing the idea structure, representing a central term or idea, are helpful in obtaining a gestalt image of the topic. As example, Figure 1 shows the terms linked with the central term – caries – for the period, 1980-2013.

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

****

The ideas are divided into dimensions representing the various components making up the meaning of caries. The dimensions providing the larger number of ideas were disease and outcome. The treatment dimension is relatively small and doesn’t include the predominant corrective procedure suggesting that approach is no longer questioned as part of research studies.

**Selection of Vocabulary:** The selection of informative terms is based on the authors’ use of these. Four measures are considered. The first is the ***grammatical class*** of the term. Informative terms are members of the nouns, adjectives, or gerunds classes. The second is the ***frequency of use of the idea***. The ideas used most often tend to represent a consensus acceptance of relevance. The third is the ***consistency of use*** across different situations. That depicts another form of consensus wherein the term is combined with other relevant terms to describe a variety of situations. The fourth is the ***specificity of the terms*** in the ideas. The terms that are more specific to the topic tend to be involved in a larger number of ideas within the involved sentences.

**Verification of Informative Terms and Ideas:** The selection of informative terms can be evaluated by comparing the terms selected by the software with those used by the subject specialist-authors. This is an important part of the quality-control process. By comparing terms captured by the two methods, the performance of the software can be determined. In the sentences that follow, terms are color coded to represent different situations. The terms highlighted in red indicate that the software properly selected them. Those in blue denote that the software missed them. Terms highlighted in blue were used by the authors, but, were not included in the vocabulary or not properly identified.

Exhibit 1 shows examples of the verification process using a document published in 2010 and two from 2013. The format is as follows:

1. The source data were given including the identification number assigned by PubMed (PMID). That number makes retrieval rapid and accurate.
2. The involved sentence was shown with the informative terms highlighted. Each term is identified once within a sentence.

Behind each sentence is a ratio giving the number of terms identified by the software relative to the number of terms available for identification.

# Exhibit 1. Sentences and Verification of Informative Term Capture by Software.

***Source:*** [***Yasuda Y***](http://www.ncbi.nlm.nih.gov/pubmed/?term=Yasuda%20Y%5BAuthor%5D&cauthor=true&cauthor_uid=21174743)***1,*** [***Tatokoro M***](http://www.ncbi.nlm.nih.gov/pubmed/?term=Tatokoro%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21174743)***,*** [***Yokoyama M***](http://www.ncbi.nlm.nih.gov/pubmed/?term=Yokoyama%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21174743)***,*** [***Koga F***](http://www.ncbi.nlm.nih.gov/pubmed/?term=Koga%20F%5BAuthor%5D&cauthor=true&cauthor_uid=21174743)***,*** [***Saito K***](http://www.ncbi.nlm.nih.gov/pubmed/?term=Saito%20K%5BAuthor%5D&cauthor=true&cauthor_uid=21174743)***,*** [***Masuda H***](http://www.ncbi.nlm.nih.gov/pubmed/?term=Masuda%20H%5BAuthor%5D&cauthor=true&cauthor_uid=21174743)***,*** [***Fujii Y***](http://www.ncbi.nlm.nih.gov/pubmed/?term=Fujii%20Y%5BAuthor%5D&cauthor=true&cauthor_uid=21174743)***,*** [***Kawakami S***](http://www.ncbi.nlm.nih.gov/pubmed/?term=Kawakami%20S%5BAuthor%5D&cauthor=true&cauthor_uid=21174743)***,*** [***Kihara K***](http://www.ncbi.nlm.nih.gov/pubmed/?term=Kihara%20K%5BAuthor%5D&cauthor=true&cauthor_uid=21174743)***. [Successful long-term management of hepatic and lymph nodes metastases of ureteral cancer by multimodal treatment including radiofrequency ablation].*** [***Nihon Hinyokika Gakkai Zasshi.***](http://www.ncbi.nlm.nih.gov/pubmed/?term=21174743) ***2010 Nov;101(7):758-63. PMID 21174743***

***Sentence 5: After three courses of combination chemotherapy consisting of gemcitabine and cisplatin (GC), one tumor completely disappeared and another achieved a partial response. (4/5)***

# *Source:* [*Katsoulis J*](http://www.ncbi.nlm.nih.gov/pubmed/?term=Katsoulis%20J%5BAuthor%5D&cauthor=true&cauthor_uid=21435159)*1,* [*Fierz J*](http://www.ncbi.nlm.nih.gov/pubmed/?term=Fierz%20J%5BAuthor%5D&cauthor=true&cauthor_uid=21435159)*,* [*Iizuka T*](http://www.ncbi.nlm.nih.gov/pubmed/?term=Iizuka%20T%5BAuthor%5D&cauthor=true&cauthor_uid=21435159)*,* [*Mericske-Stern R*](http://www.ncbi.nlm.nih.gov/pubmed/?term=Mericske-Stern%20R%5BAuthor%5D&cauthor=true&cauthor_uid=21435159)*. Prosthetic rehabilitation, implant survival and quality of life 2 to 5 years after resection of oral tumors.* [*Clin Implant Dent Relat Res.*](http://www.ncbi.nlm.nih.gov/pubmed/?term=21435159) *2013 Feb;15(1):64-72. doi: 10.1111/j.1708-8208.2010.00328.x. Epub 2011 Mar 22. PMID 21435159*

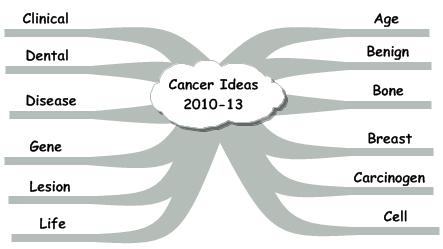
***Sentence 1: After oral tumor resection, structural and functional rehabilitation by means of dental prostheses is complex, and positive treatment outcome is not always predictable. (7/7)***

***Source:*** [***van der Waal I***](http://www.ncbi.nlm.nih.gov/pubmed/?term=van%20der%20Waal%20I%5BAuthor%5D&cauthor=true&cauthor_uid=23229266)***1. Are we able to reduce the mortality and morbidity of oral cancer; some considerations.*** [***Med Oral Patol Oral Cir Bucal.***](http://www.ncbi.nlm.nih.gov/pubmed/?term=23229266) ***2013 Jan 1;18(1):e33-7. PMID 23229266***

***Sentence 8: Reported reasons for patients' delay range from fear of a diagnosis of cancer, limited accessibility of primary health care, to unawareness of the possibility of malignant oral diseases. (8/8)***

**Idea Maps:** Idea maps can take various formats. These include:

1. Alphabetic array of the terms linked with a central term.
2. Frequency array of terms linked with a central term.
3. Dimensional array of terms linked with a central term.

**Figure 2. Alphabetic Array Using Informative Terms From Cancer Related Ideas.**

The ideas shown in Figure 2 depict relationships between the central term – cancer – and the informative terms occurring with higher frequency. The terms are arrayed beginning on the right and continuing to the left. The alphabetic array is useful as a type of subject index providing an assessment of the importance assigned to each idea by the subject specialists, assuming that importance is determined by frequency of use. The alphabetic array is helpful to the subject specialist interested in a specific idea by highlighting its existence and frequency.

**Figure 3. Frequency Array Using Informative Terms From Cancer Related Ideas.**

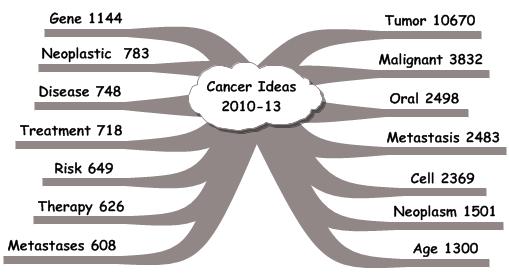


Figure 3 shows the higher frequency terms related to the central term – cancer. The terms are shown with the most frequently occurring ones on the right and proceeding down and around on the left. This display serves as an index showing the ideas considered in terms of relevance (i.e., frequency) by the subject specialists. This display is most useful to the student interested in acquiring an understanding of a new topic. By focusing on the highest frequency ideas, the student can rapidly prepare a description of the topic.

Figure 4 shows the higher frequency ideas involving the central term – cancer – arranged in dimensions representing major components of the topic. Those dimensions are shown as arms in the idea map. The terms were classified using the meaning and/or function. This classification scheme is useful in formulating research designs. The dimensions represent components that must be addressed. The personal factors describe the characteristics and attributes of the individuals or materials chosen to study. The environmental factors describe the physical, social, and psychological characteristics of the environment in which the individuals live. The subject factors represent the physical, chemical, social and psychological attributes that make up the body of knowledge regarding the topic. The intervention factors are physical, chemical, biological, psychological, or educational acts used to influence change in specific subject factors. The outcome factors represent the characteristics or attributes being affected by the interventions. The methods describe the procedures used in the study.

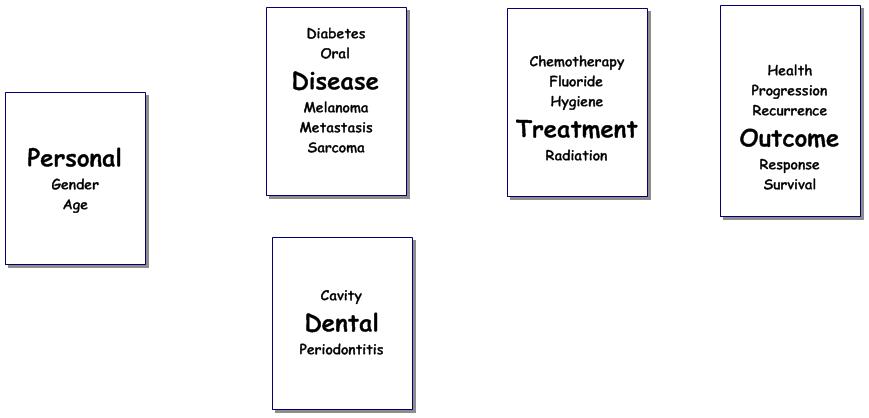
**Figure 4. Dimensional Array Using Informative Terms From Cancer Related Ideas.**

****

**Idea Maps and Research:** By identifying the ideas within each dimension, different constructs can be considered. For each, the higher frequency ideas form a basic structure of acceptable and relevant knowledge. Often, lower frequency ideas will be introduced as testable elements. The composite will determine the impact of the low frequency ideas in expanding or enhancing the knowledge structure.

Idea maps serve to provide a gestalt view of the knowledge structure based on the ideas. Figure 5 shows an expansion of the dimensional map. Each of the terms was linked with cancer. An initial question is the interrelationships among the terms.

**Figure 5. Dimensional Idea Map Including Terms Linked with Cancer and with Tumor.**

****

**Syntheses:** The first of the higher cognitive functions is synthesis. The research design template can be used to construct a number of different syntheses. Examples might be:

1. Gender 🡪Diabetes 🡪 Oral 🡪 Cancer 🡪 Periodontitis 🡪 Chemotherapy 🡪 Response
2. Age 🡪 Melanoma 🡪 Cancer 🡪 Hygiene 🡪 Recurrence
3. Metastatic 🡪 Cancer 🡪 Chemotherapy 🡪 Progression.

These examples and other syntheses illustrate a different approach. While each synthesis is a possible study, the important action is no longer constructing a study question. Instead, the question becomes – “***Which of the many possible syntheses is the most important to study?”*** This shift from searching for a study question to developing an evaluation system for prioritizing the syntheses is an example of objective conduct of the higher cognitive functions – ***comparison, evaluation, judgment, and application.***  Each function involves the development of measures, criteria, and decision-rules. This critical analysis plus the special interests and experiences of the analyst make up the critical and creating thinking process. With the ideas as the essential data, the transparency involved in performing this procedure is enhanced.

**Initial Uniformity of Knowledge:** Given the existence of the idea database, uniformity of knowledge can be realized at least at the outset. With a comprehensive and accurate representation of the subject specialists’ ideas in a usable form, each investigator starts with an accepted and comprehensive set of specific ideas, concepts, and issues. What is done subsequently is a matter of individual intellectual prowess. Working with ideas offers the opportunity to make those subsequent cognitive functions more transparent and subject to quality-control. In that way, uniformity of approach to critical and creative thinking is a possibility. The findings can be diverse depending on the measures, criteria, and decision-rules employed. The uniformity stems from the application of the specific procedures in carrying out each cognitive function.

**Uniformity of Study Design:** Without the centralized idea database, coordination of research required at least two levels of oversight. One associated with funding and the second with the directions of research. With the centralized idea database, the changing idea map serves to coordinate the research by displaying the work accomplished and enabling investigators to identify the elements needed. Communication among the investigators via email and at conferences could fulfill the need for interaction and assignment of future work. In this milieu, funding of the preliminary studies is not only acceptable but important in rapidly describing and developing definitive studies.

**Communication Among Investigators:** The ability to work in the cooperative group setting while at the same time performing vital preliminary descriptive-intervention studies is important in expanding the knowledge regarding a particular topic and in benefiting the individuals effected by the problem. The central database of ideas would assist in bringing a consensus among the experts with respect to the starting level of knowledge and the need for coordinated investigations. The intellectual benefit would be seen in a transparent, quality-controlled approach to critical and creative thinking.

The idea database is a feasible approach to the important first steps in the organization and utilization of information:

1. The process of building the idea database is accurate and comprehensive. Of importance, the quality control process enables determination of the degree of accuracy and the extent of comprehensiveness.
2. The resulting ideas are relevant to the topic. The authors demonstrate their intentions relative to the topic by linking the topic descriptor to related terms.
3. The time and effort involved is shorter than other available approaches. The analysis of the text and subsequent organization takes about 0.3 minutes for each 250 word document.
4. The associated costs are considerably cheaper as the software is performing the bulk of the effort. The cost of the individual’s time and energy is transferred from the clerical tasks involved in retrieval to the intellectual ones.

**Research Agenda:** Once the idea database is operational, the development of a research agenda is efficient because of the ease in arranging the ideas to form different syntheses. The higher frequency ideas provide a consensus description of what’s known at a given time. The lower frequency ideas represent two situations – one depicting the disappearance of ideas with time. The second represents the emergence of new ideas. Temporal analysis can readily identify the two situations. The impact of the new ideas on the established structure can be assessed and those likely to represent important change can be considered in new research strategies.

Since ideas are independent building blocks they can be combined in various ways depending on the intent of the analyst either to duplicate existing idea structures or to build new ones. The central database and the higher cognitive function algorithms create a different type of expert. Traditionally, a specialist would be chosen for his/her wisdom and opinions. With the tools described above, the expert is needed to clarify and expand the higher cognitive functions while the more clerical/mechanical ones are converted to a transparent, evidence-based system. By making the analysis transparent and quality-controlled, the needed documentation is readily available. This facilitates the shift to true intellectual prowess by the expert.

In a similar fashion, the student, new to the subject, can begin learning by solving problems and by building new idea structures. The ability to acquire, organize, and utilize the ideas enhances the transformation from novice to professional. The need to spend long hours in the library stacks is replaced by a need to spend time thinking and researching. Those actions could yield an operational description of critical and creative thinking.

With the ideas organized and usable, students can build numerous idea structures representing a given topic. This array shifts the focus from a single answer to a spectrum of possibilities each with desirable and undesirable characteristics. Weighing those attributes and developing rationales is an example of the critical thinking process. Translating those functions to transparent, quality-control procedures is an example of the Idea Analysis approach.

**Transforming Ideas to a Research Design:** Ideas can be explored by considering how the authors used them in sentences. Authors select and arrange the informative terms to express the ideas of interest in a coherent way. The arrangements are not the only ones possible. Since ideas are independent building blocks they can be combined in various ways depending on the intent of the analyst.

**Multi-Dimensional Research:** An important feature of research involving multi-dimensional knowledge is the realization that numerous subject factors and interventions may be required before significant change is observed in an outcome factor. This form of research differs from the traditional laboratory experiment where a single factor is influenced to accomplish a specific change. The complexity of multi-dimensional research requires that learning progresses in phases.

1. The first phase deals with identification of possibly ***relevant attributes*** of characteristics.
2. The second phase deals with determination of ***possible changes*** in those attributes.
3. The third phase deals with comparison of ***degree of change***.
4. The fourth phase deals with ***short term and long term effects*** of the change.

**Combinational Designs:** Frequently, there is an attempt to combine those phases to save time. The accepted combination is an interventional study requiring large samples of study subjects, comparable interventions designed to show the influence of the new one, and longer periods of observation. Those studies could be expensive involving numbers of personnel, more measurements, and questionable results. Without the proper problem descriptors, the success of the project is unclear.

**The Descriptive-Interventional Design:** The distinction between descriptive and interventional studies is in the acceptance of current knowledge from the literature. Descriptive studies are based on the assumption that the evidence regarding a phenomenon is weak and more is needed. Intervention studies are based on the assumption that the current description is adequate to develop a test of effectiveness. As a result of the investigation, the description is enhanced and the evidence dealing with the relationships is increased. Intervention studies are of value in building an evidence-based, data-driven description of the phenomenon.

Researchers often adopt descriptive studies instead of interventional ones. The descriptive ones were intended to develop the knowledge structure representing the topic. Often, these studies took on a life of their own, and became the end rather than the starting point originally intended.

An alternative to this emphasis on description is the preliminary interventional study. This design combines the descriptive elements with an intervention. The sample size is relatively small and the measurement set restricted. The objective of the study is to develop a ***more specific description together with an initial assessment of the degree of change*** that might be observed. The results of this type of study could lead to other small descriptive-interventional studies. The findings would eventually be sufficient to formulate a definitive study with the essential subject factors and interventions in place. That study would conform to the requirements of sample size, personnel, conduct, and duration. The chances of developing interpretable results would be enhanced.

This descriptive-interventional study design often is described as preliminary or pilot. The study is effective in developing more specific questions for study but rarely produces answers. While these preliminary studies provide insights with respect to the characteristics needed for accurate description plus estimates of the degree of change offered by the interventions, funding agencies tend not to fund them. In terms of fundable studies, there are two primary types – the large survey (descriptive) study and the large intervention study with randomized assignment to the comparable interventions.

**Advantages of the Descriptive-Interventional Design:** The advantages of this study process are numerous. The descriptive-interventional study process can be carried out in many environments providing insights across different study groups and physical settings. A measure of ‘truth’ in science is the ability to replicate the findings in different situations. This consistency may be of greater interpretability than a single large study.

A second advantage may be the identification of different relevant attributes and characteristics together with the measurements/observations depicting them. Having different investigators pick, independently, the measures, criteria, and decision-rules in each study offers a more complete picture of the knowledge structure than if everyone employs the same study method, particularly early on. Uniformity is appropriate when the definitive study is launched. The preliminary studies should employ different methods in order to determine the information provided by each.

A third advantage is the ability to employ a variety of interventions. These findings would be relevant in determining the better ones to employ in the definitive study. The combination of subject factor to be influenced, the intervention chosen to be the influencer, and the outcome factor representing the degree of change provide an important component in the determination of the ultimate study. Having different investigators determine these interventions again offers a form of consensus when the same intervention is studied in different places and a diversity of results when different ones are studied.

**Essential Difference – Now vs. Proposed:**  Small studies have been performed through time and the results published. The uncertainty in this process lies with the question of use of the information. Since these studies are small in size and may be unable to declare definitive findings, the attention they receive by other investigators could be variable. In contrast, if these small descriptive-interventional studies are part of a coordinated approach, the findings are added to the knowledge base and used to develop new studies. Progress is coordinated and knowledge acquisition accelerated.

**Idea Maps, Research Studies and Uniformity:** The advantages of uniformity are important in science and learning. However, the emphasis on uniformity has been misplaced. The importance of a uniform knowledge base of data describing the topics cannot be dismissed. Previously, the uniformity was seen in the development of bibliographic repositories, containing documents and records describing the presumed content of those documents. This warehousing approach was a valuable first step. In numerical problems, the concept was changed to a combination of warehousing and knowledge base offering credible data available via a number of portals. This combination has not yet been accomplished with text and is the objective of the idea analysis approach.