**Chapter 1. New Tools to meet New Challenges**

**Abstract**

**In this discussion, the definition of an idea is divorced from the personal and private creative act and is considered simply as a fact provided by an author-specialist. This offers a dramatic change in use. Software can recognize the existence of an idea and present it without inferring importance. This separation of recognition and relevance offers the ability to use the ideas in building descriptions of a topic as well as in building new research strategies. This enhanced capability involves a significant reduction in overall time as well as a shift from the traditional 90% perspiration to 90% inspiration. In addition to the savings in time and reduction in tedious acts, use of computer-supported algorithms offers the opportunity to construct intellectual paths approximating critical and creative thinking and to make these more transparent. This methodologic approach to continual learning could result in more rapid acquisition of new knowledge.**

**Introduction**

**This discussion reports the use of *ideas* in stimulating efficient processing of the ever-increasing volume of text and in stimulating a form of transparent creativity. The past literature is a treasure house of ideas with a median of 8-10 per sentence. The challenge is to capture those ideas and organize them so that they can be used for description of past as well as future.**

**This flood of ideas (literally millions) poses a challenge to efficient processing. To meet this, tools representing the most effective use of computer technology are required, or are they? Is it possible that tools previously used in processing numerical data provide the basis for effective handling of alphanumeric data? This discussion explores that possibility and shows that such tools can be used to develop new descriptions of topics and the construction of new research strategies. To accomplish this, three tasks are required –**

1. **To identify, extract, and organize the ideas as a data resource.**
2. **To use the ideas as a portal for retrieval of original documents.**
3. **To use the ideas as a means to identifying new organizations leading to new research strategies.**

**Following the World War II, there was a surge of information that ranged from scientific subjects to a wide variety of others. This increase was described as exponential. Interestingly, the escalation in volume didn’t stop but instead increased in subsequent years so that a typical processing task now involves thousands to millions of documents. Using traditional paper-oriented procedures (computer supported or manual) resulted in delays of months to years in identifying, organizing, evaluating, and selecting the information to be presented as a description of a topic. The time spent in developing specialty interest in a subject significantly increased with a designated block of time spent simply trying to make sense of the volume of information available. The accepted practice of using original documents became less feasible and summaries in either book form or journal produced reviews became more useful. This reliance on third-party summaries increased in spite of the larger time gap between the presented material and current reports.**

**To further complicate the situation, new or potential authorities from numerous countries contributed to the body of information in a topic. These non-traditional ‘experts’ literally flooded the subject, making their contributions relevant simply by number of published documents. As research and its quest for new knowledge shrunk in the United States and Europe, the resulting vacuum was filled by manuscripts from South America, the Middle and Far East. As such, the established behavior of citing and quoting eminent authorities required altering because of failure to recognize who was and wasn’t an authority.**

**Computerized search engines became more sophisticated identifying millions of documents or websites in seconds. The claim was that computer support had favorably affected information processing. However, the problem of determining the content of each document or website relied on the traditional paper-oriented approaches. These involved identifying key words (i.e., terms that conveyed information regarding the topic) or highlighting phrases contained within a document. Organizations of these informative elements were intended to describe the contents of the documents.**

**Another advance in understanding the content of a document involved the linking of two or more informative terms. These *hyperlinks* were intended to augment the meaning of the document by providing informative phrases. These identification and organizational methods involved the use of the impressive pattern-recognition capability of the computer.**

**Linking informative terms was a significantly potential breakthrough in content analysis of text. It had advantages:**

**1. The author’s vocabulary was used. Assuming that the author was a specialist in the subject, his/her choice of informative terms should be more specific in describing the content of the document.**

**2. The author’s choice of words used in a linking should offer greater insight into the topic as compared with arbitrary pairings made by an individual who may not have awareness of the insights expressed in the document being processed. Search engines, until recently, did not reveal the contents of the documents held in the repository. The current presentation is based on the analyst’s choice of the first term. The engine displays linkages involving that term with others from the repository. These linkages may be combinations provided by the authors or arbitrary ones formed by the engine. Distinction between the two is not possible.**

**The problem with linking is not the concept of pairing but the domain considered in identifying pairings. Consider the linking of two terms in various domains:**

**1. First and last page of the document.**

**2. Top and bottom of the same page.**

**3. First and last sentence of a paragraph.**

**4. Two terms within the same sentence.**

**The author is responsible for choosing the informative terms and in expanding on their meaning by including them within close proximity. The simple sentence illustrates this. The sentence consists of three terms – a subject, a verb, and an object. The subject and object form a thought (an idea) while the verb denotes the action or state of being considered by the author in providing the idea. Interestingly, the idea is unchanged irrespective of the change of verb. As such, the idea is an important element in conveying meaning and has been recognized as such for centuries. Links involving terms in contextually different areas of the document would have less accurate meaning than two terms within the same sentence.**

**The problem with an idea is its definition. An idea is defined as a product of subconscious and conscious mental activity. That definition implies that an idea is a private and personal thing. The creative process involved cannot be communicated and may be unrecognized by even the creator. The idea creation has been described as “Eureka!” or “Wow!” implying a sudden dramatic awareness of the new idea.**

**If the creative nature of the idea is separated from the task of identifying ideas, a radical change in text processing is possible. Assume that the definition of an idea *is two or more informative terms coupled by the author and presented within the domain of a sentence.* This definition makes no statement regarding the creation of the idea. Instead, it states that an idea, created by an author, in any way possible, can be identified by the presence of the involved words in a sentence. This identification process can be translated into an algorithm for use by a computer. The ability of the computer to recognize patterns can be harnessed to identify the informative terms used by the author in each sentence. The editing capabilities of the computer can be employed to combine the informative terms into pairs and add bibliographic data indicating the document and sentence containing the pair. These records can be combined and stored in a data file.**

**This process can be performed tirelessly by software and at remarkable speed with exceptional accuracy. As such, the time required to determine the content of a document can be reduced to fractions of a minute\.**

**The resulting database of ideas serves two important functions:**

**1, Precise identification and retrieval of documents. With the ideas extracted and organized, the analyst can select a relevant idea and retrieve the document(s) containing that idea. The involved sentences can be organized as a first draft of a new description with the assurance that accuracy and completeness are maintained. The resulting audit trail of procedures transforms the private and personal process of selecting text to a transparent one easily quality-controlled.**

**2. The ideas are independent of the text in which they were originally found. As such, these ideas can be combined in various ways to consider alternative descriptions and potential research strategies. This transparent process approximates *critical thinking* and *creativity* by enhancing the production of *measures* describing the phenomena, *criteria* for using those measures, and *decision-rules* for determining results. By employing software algorithms, these procedures can be performed with quality control and public awareness of the intellectual activities involved. The focus on the ideas eliminates the traditional reliance on the quotation-citation reference process (i.e., *who said, what was said)* enabling a concerted focus on development of new knowledge.**

**The hyperlinking of key words plays a significant role in the capture of ideas presented by the author-specialists. The focus on the sentence as the relevant domain for hyperlinking ensures that each idea will be contextually correct whether or not the idea was considered important by the presenting author. Insignificant ideas at one time have the potential to be powerful in another. As such, the capture and organization of the total set of ideas from each document offers an array of possibilities that could result in new information. This array is not unlike a distribution of values in a statistical problem. The majority of such may be comparable in describing a single situation while a few may suggest a different one. The ability to formally recognize the difference is an important contribution from statistical analysis. In a similar fashion, the ability to formally recognize an array of ideas leading to new insights would be considered an important advance.**

**The development of tools capable of formally and transparently recognizing the arrays of ideas and differentiating the new potentials would be a worthy accomplishment. Toward that end, the investigation using ideas as the essential information has been ongoing since the 1970s. The approach has been tested and found to be feasible and practical against a wide range of text describing different subjects and prepared by a cadre of authors embracing different training, orientations, and writing styles. The process is based on the realization that each author and each language uses the sentence as the conveyor of ideas. By capturing those ideas and organizing them as a database, the knowledge of those specialists can be shared. In doing so, algorithms can be employed making the process transparent. The challenge is to develop those algorithms.**

**Algorithms Used in Building the Idea Database.**

**A fundamental question employed in the design of the research was – *Could the ordinary and well-accepted editing procedures used in processing numeric data be translated to deal with text?* If so, a number of algorithms could be constructed for testing.**

**The search engine, developed in conjunction with the bibliographic repository containing the digital versions of the original documents, was an important tool already in existence for dealing with text. This engine provided access to the specific documents that dealt with a subject. This identification process began with key terms and evolved to include designated phrases. Boolean operations (AND, OR) were employed to combine separate terms to form linkages that may or may not be present in the actual text of the documents. The intended use of the search engine was to identify the specific documents containing the search phrase and thus, provide precisely the information desired by the user. By developing a search statement intended to capture all of the documents describing a subject, the stage would be set for the automated text analysis leading to the idea database.**

**Unfortunately, search engines have design flaws which result in failure to capture the specific subject and eliminate others. A possible reason may be the multitude of subjects presented within a document by the authors. That is, documents do not necessarily describe one topic and exclude all others. As such, selection may result in a proportional array of relevant and irrelevant documents depending on how the individual author used the search combination in his/her document. Authors tend to include a median of 8-10 ideas within a scholarly sentence and in turn may include a number of subjects within a document. This conglomerate of ideas and topics may be simply a reflection of the complexity enjoyed by a subject as it is considered by different disciplines and specialties.**

**Idea Identification and Organization.**

**Given a set of documents selected, using a search statement, software may be constructed to perform the idea identification process. The tasks involved are:**

1. ***Separation of individual sentences using ending punctuation* (periods, question marks, etc.)**
2. ***Identification of informative terms (nouns, adjectives, or gerunds) using endings.*  Nouns may end in suffixes such as –ion. Adjectives may end in –al or –ive. Gerunds end in –ing and perform a descriptive role rather than the typical verb action.**
3. ***Combinations of informative terms can be accomplished in a brute fashion resulting in defined ideas, each consisting of a pair of terms.***

**Each idea is assigned bibliographic data (document membership, year published, and sentence membership). These data allow precise retrieval of the document containing the idea of interest.**

**The third set of tools in building the idea database was incorporated in the Excel database. These included identifying, copying, and sorting. Namely the typical operations involved in editing numerical data and arranging it for use. This set of operations in the database was readily expanded to deal with alphanumeric data. Excel was chosen as the final repository format because of the widespread use of the software for a significant period of time.**

**Algorithms Used with the Established Idea Database.**

**The Excel idea database can be sorted using the Primary Term as the sort column. The result is an array of idea records. Each deals with an informative term and its related terms. The subtotal function in the Data operation can be used to determine the number of ideas involving each selected term. This distribution of idea frequencies can be formalized as a data table.**

**Ideas as Independent Building Blocks.**

**An important advantage in using ideas is the ability to build different arrangements of the ideas. A template from clinical trial research can be used to construct different arrays. Given that set of possibilities, the challenge is to develop *measures* describing the attributes of each array. These measures can be used together with *criteria* for management. Outcome (i.e., selection or discard) can be accomplished using developed *decision rules.* The construction process is a clerical operation accomplishable by either manual or computer means. The development of measures, criteria, and decision rules, however, is a real intellectual function.**

**Research Template.**

**The template is composed of 6 dimensions. Each contributes a significant component of the topic. The dimensions are:**

1. ***Personal Factors* – those measures (attributes) that describe the individuals or elements being studied. Typical ones would be gender, age, height, or weight.**
2. ***Environmental Factors* – those measures that describe the locale in which the study subjects exist. Factors such as physical, social, psychological or political variables are relevant.**
3. ***Subject Factors* – variables that describe the topic. These include clinical, physical, laboratory, and psychological.**
4. ***Intervention Factors* – variables that describe intent to alter subject factors in some way. Interventions include physical, chemical, pharmacological, educational, psychological, political, and social.**
5. ***Outcome Factors* – variables that describe the result of the intervention on the specific subject factor(s). These include response, survival, quality of life, and failures of various types.**
6. ***Methods* – measures that describe the methods used in studying any of the other dimensions.**

**Types of Resulting Designs.**

**By selecting specific dimensions to populate, the ideas chosen formulate specific study designs intended to answer different questions. Examples of these designs would be:**

1. ***Relational Designs*. These are descriptive. Examples would include:**
2. **Personal 🡪 Environmental.**
3. **Personal 🡪 Subject Factors.**
4. **Personal 🡪 Outcome Factors.**
5. **Personal 🡪 Personal.**
6. **Environmental 🡪 Environmental.**
7. **Environmental 🡪 Subject Factor.**
8. **Etc.**

**Each of these designs is intended to link the factors in one of the descriptive dimensions with measures in the other dimension in order to describe the relationship(s) involved. These designs seek to identify pertinent measures that possibly influence any interventions attempted without actually including an intervention in the design. Typical examples of the relationship design would be the *survey* whether done using individuals from some population or tissues in some laboratory setting. In each situation, the intent is to determine a link between variables.**

1. ***Interventional Designs.* These designs involve the intent to alter some subject factor by imposing a condition (treatment) on that factor and then determining the result of that effort. The full-blown interventional design involves all 6 dimensions. Study subjects are selected with particular attention to Personal and Environmental characteristics. Specific Subject factors are designated as relevant for change. The Treatment program is selected to attempt to accomplish the change. The change characteristics are summarized by specific Outcome factors. Finally, Methods pertinent to each dimension are selected so that the process is as effective as possible.**

**The advantages associated with this design are two-fold. First, insights into the action-reaction of the intervention🡪outcome link can be determined. If the intent is to correct some problem (illness, etc.) the results from the intervention-outcome could lead to new approaches and resolution of the problem. Second, insights into the relationships between personal, environmental, and subject factors could be more informative. Determining the important characteristics involved in designing study of an intervention-outcome could be as relevant as the result of the intervention.**

**An additional advantage could be seen in the selection of the methods employed in the study. The need for accurate, efficient methods is obvious. However, differentiation among methods may be more difficult without the benefits of the intervention.**

**Criteria for Assessing Design.**

**A primary issue would be the inclusion of dimensions in the study design. That is, does each dimension contribute important information in understanding the overall intent of the design? Prior studies provide insights into the relevance of Personal, Environmental, and Subject factors. If those studies are lacking, preliminary relationship or interventional designs would be useful in contributing the evidence needed.**

**Issues related to intervention may be more specific. In developing an intervention, measures such as dose, method of delivery, frequency of delivery, and likely responses would be helpful in assessing the potential merits of the intervention. The outcome factors should clarify the possibility of favorable or detrimental results, the degree of response, the length of the response, and the conditions associated with relapse. If the intervention is ineffective, measures such as -- the time to definitive failure, the complications associated with failure, and possible corrective measures – could be beneficial.**

**These criteria could be scored using a simple binomial scale or more elaborately using a Likert scale. Once scored, the summed value for each developed design could be used to develop rankings of the ‘best’ to the ‘least’.**

**Decision Rules.**

**The rankings could be used as a starting point for final evaluation. Depending on the distribution of the ranks, several subgroups of designs could become evident. Focusing on the top ranking, a type of Delphi process could be employed. Additional measures of the characteristics of each design could be considered, scored, and ranked. This iterative process could result in one design that offers more advantages than close alternatives. If so, that design might be selected for translation into a definitive protocol.**

**The Value of the Protocol.**

**Translating a design into a protocol involves expanding the shorthand notes for each dimension into full blown sentences and procedures. In doing that, defects – gaps or inconsistencies – may be evident, changing the assignment of Delphi scores and requiring selection of an alternative design. Once the selected design is translated to a working protocol, the actual execution is next.**

**Transparency and the Creative Process.**

**Much of what is suggested above is accomplished by the experienced investigator without necessarily recording each step. The advantage of the transparent and ‘formal’ approach described is evident. Third-parties can benefit from the thought process employed by the investigator. Quality control procedures are possible ensuring that the deliberation process is as accurate and efficient as possible. More importantly, errors in the intellectual path may be more readily identified before time, energy, and resources are wasted. The time involved is considerably shortened. Using paper-oriented methods, significant periods of time are spent in essentially clerical operations. Using computer-supported algorithms, the time spent in intellectual functions (measures, criteria, decision-rules) is increased and the clerical is performed by the software.**

**The Role of Ideas.**

**Ideas serve as an engine in developing transparent designs. The frequency of authors’ use of specific ideas serves as a type of consensus reflecting ‘importance’. These ideas often serve as the framework representing the topic. Less frequent ideas that appear later in time may represent the springboards for new information and development of new knowledge. Ideas fit easily into the design template by being linked with terms in the same or other dimensions. By classifying each term in each idea into its member dimension, populating the template is a matter of minutes. Classification into dimensions often is arbitrary and is based on the more frequently used meaning of the term. That offers the advantage of considering the different dimensions possible for an idea. Those variations may change the intent of the design offering different situations to study. If so, the array of study designs is rapidly constructed and can be scored as described.**

**Summary**

**The well-established definition of the idea focused on its creation. The definition correctly cited the subconscious and conscious acts involved in generating an idea. In addition, the product of this creative function could be described by a single term, a phrase, or a complete text. This somewhat mystical and vague description of an idea meant difficulties in passing the ideas and their value from mentor to student. That process, if accomplished, would take months to years.**

**This report considered the value of an idea in modern text processing. With the deluge of information available in virtually every topic – scientific or other – the emphasis is on more rapid learning. Labels such as problem-based learning, just-in-time learning, guided learning, and informed learning describe instruction focused on student identification of issues representing a topic. These forms of instruction use the teacher as a mentor-coach-guide.**

**The subject matter considered is reduced in one way or another. This reduction is a requirement of the paper-oriented approach. Ideally, the millions would be reduced to one before re-expansion based on that idea.**

**The Contextual Analysis approach differs by considering the various arrays of ideas representing different study designs. These are scored and ranked in terms of potential success. The selection of a favored design doesn’t mean that all others were bad. Instead, selection is tied to the interests and perceptions of the individual. In an ideal cooperative organization, different individuals would be expected to select different designs thus expanding the potential for learning new knowledge.**

**This arrangement is common place in new treatment development. There are physicians who specialize in determining the likely dose of a new agent to be administered. Others determine the beneficial and detrimental effects of the new with or without other agents. Still others determine the long term effects of the new treatment regimen. Finally, others evaluate the effects of the new regimen in expanded use.**

**The modern definition of an idea – *two or more informative terms combined by the author-specialist and presented in a sentence* – offers enhanced capability in using these ideas to rapidly and accurately build new descriptions of topics or new learning (research) strategies. An important advantage associated with use of ideas is the ability to develop quality-control procedures associated with the cognitive functions involved in intellectual activities. This offers a previously missing form of transparency making learning of creative paths easier and more intuitive.**

**With the ideas and computer-supported algorithms, the ability to develop measures, criteria, and decision-rules (the elements inherent in critical and creative thinking) is facilitated. The intent of search engines is to identify documents of potential interest. The objective of Contextual Analysis is to identify essential content and its inclusion across documents. The previously emphasized relevance of the document is replaced by a new focus on document content. This shift from the total to the specific elements offers greater flexibility and is in keeping with the multifaceted content of a single document.**