Application of Context to Fast Contextually Based Spatial Authentication Utilizing the Spicule and Spatial Autocorrelation

Gregory Vert
Center for Secure Cyberspace
Louisiana State University, and
Texas A&M Central, Texas
(206) 409-1434
gvert12@csc.lsu.edu

Jean Gourd
Center for Secure Cyberspace
Computer Science
Louisiana Tech University
(318) 257-4301
jgourd@latech.edu

S.S. Iyengar
Center for Secure Cyberspace
Computer Science
Louisiana State University
(225) 578-1252
iyengar@csc.lsu.edu

ABSTRACT

This paper proposes an integrating mathematical method that unifies a new Contextual Processing model with that of the Spicule visual authentication method. In previous work the Spicule has been initially determined to be much faster at generation of authentication signatures for spatial data than standard encryption methods. It however could be much faster than it already is if a method was designed that could reduce the number of spatial objects it has to generate authentication signatures for. Previous experiments required Spicule to authenticate all spatial objects in a set for comparison against encryption methods. This paper provides brief overviews and background on the Spicule, and the new Contextual Processing model. It then proceeds to present the integrating mathematical approach of localized spatial autocorrelation. Finally an algorithm and the overall application of the method is presented by which limited sets of spatial object are mathematically selected for authentication when they are germane to a spatial query.

Keywords
spatial data authentication, contextual processing, contextual processing security, spatial autocorrelation.

1. INTRODUCTION

Contextual Processing (CP) has been around the research fields of Computer Science off and on for years. It has always had limited and focused application. However as the world has faced such events as 9/11, Indian ocean Tsunami, and Three Mile Island nuclear disaster there has been a need for more advanced processing paradigms especially ones that consider spatiality and temporality.

The goal of research in this area has been to link the environment a machine exists in to how the machine may process information. An example typically given is that a cell phone will sense that its owner is in a meeting and send incoming calls to voicemail as a result. Application of this idea has been applied to robotics and to business process management [1].

Some preliminary work has been done in the mid 90's. Schilit was one of the first researchers to coin the term context-awareness [2,3]. Dey extended the notion of a context with that of the idea that information could be used to characterize a situation and thus could be responded to [4]. In the recent past more powerful models of contextual processing have been developed in which users are more involved [5]. Most current and previous research has still largely been focused on development of models for sensing devices [6] and not contexts for information processing.

In addition to CP, there has also been an explosion in the amount of spatial data being generated and a heavy reliance on such data. Considering the use of GPS and Google Earth, this type of data is not the alpha numeric types of information that has been traditionally managed. One characteristic of the data is that it is voluminous and has spatial relationships inherently that much be preserved in its management. Coincidental with this fact, has been an increasing need to secure such data as it is transmitted across the internet. Traditional authentication methods have relied on dated concepts of hashing and encryption which are computationally impractical on large volumes of data. Instead of building faster processors the performance bottleneck of authentication can be addressed by *working smarter, only do what is required and ignore the rest of the noise*.

The following sections provide overviews on two brand new paradigms, that of the new CP model and the other of a visual algebra, the Spicule, that can be used for

authentication. Integration of the paradigms provides a potential path towards working more intelligently and quickly on authenticating and securing spatial information.

2. CONTEXTUAL PROCESSING

2.1 Overview

The initial development of the new CP model was based on examination of the natural disasters of the Indian Ocean tsunami, three mile island nuclear plant and 9/11. A goal was to determine what elements could be used to categorize these events. After analysis it was realized that all of them had the following categorical properties, which are referred to as the *dimensions* of a context in the model. They are:

time – the span of time and characterization of time for an event

space – the spatial dimension

impact – the relative degree of the effect of the event on surrounding events

similarity – the amount by which events could be classified as being related or not related.

Each one of the dimensions can be attributed which can be used to derive the semantic processing rules. These dimensions were discovered to be critical in the derivation of knowledge about an event because they affected the process of reasoning about an event. For instance, the time space dimensions can be utilized to reason that a tsunami in the middle of a large ocean may not have the *impact* or *similarity* to that of one just off the coast of Thailand and therefore the processing and dissemination of that information will be different. The reasoning is based in this case on the context defined by the dimensions.

The time and space dimension context driven processing will have the factors of geospatial and temporal elements to them. The geospatial domain can mean that information is collected and stored at a distance from where it may be processed and used in decision support as well as a description of the region that a context may pertain to. This means that context based information processing (CBIP) processing must have a comprehensive model to route information based on semantic content to the appropriate processing location and dissemination channels. CBIP processing can and often does have a temporal component. It can be collected over periods at regular or irregular intervals (the attribution of the dimension) and the time that the information is collected also may determine where the information is sent and the context of how the information is processed. For instance information that is collected as simply monitoring information may in the case of the Tsunami flow to research institutions around the world for storage and analysis at some point in the future. Whereas, noticing earthquakes on the ocean floor may route collected information to countries surrounding an ocean for immediate high speed analysis, critical real time decision making and rapid dissemination. Some factors that should be considered in CBIP processing are referred to as information criticality factors (ICF). These factors are further developed in ongoing research but are primarily used to drive processing decision making. They may include such attribution among other attributes as:

- time period of information collection
- criticality of importance,
- impact e.g. financial data and cost to humans
- ancillary damage
- spatial extent
- spatial proximity to population centers

These factors and many others in the model could be used to evaluate threat, damage, and criticality of operational analysis. Other factors affecting CBI processing might be based on the *quality of the data* such as:

- currency, how recently was the data collected, is the data stale and smells bad
- ambiguity, when things are not clear cut – e.g. does a degree rise in water temperature really mean global warming
- contradiction, what does it really mean when conflicting information comes in different sources
- truth, how do we know this is really the truth and not an aberration
- confidence that we have the truth

From the initial analysis of the facts describing the Indian Ocean Tsunami factors were defined that could define events and the context surrounding the event. These dimensions where defined to be the following:

temporality – defined to be the time period that the event unfolded over from initiation to conclusion

damage – the relative damage of the event both in terms of casualties, and monetary loss

spatial impact – defined to be the spatial extent, regionally that the event occurs over.

policy impact – directly driving the development of IA (security) policy both within a country and among

countries. This directly led to the evolution of security policy driving implementation because of the event.

2.2 Defining a Context

After the above dimensions were defined, the next phase of the research was to determine more rigorously how these factors might be defined and manipulated in an abstract sense. The following model component was developed where feature vectors could be utilized to define context and the factors of context. In its simplest form, a context is composed of a feature vector

$$F_n<a_1,..a_n>$$

where the attributes of the vector can be of any data type describing the event. This means that the vector can be composed of images, audio, alpha-numeric etc. Feature vectors can be aggregated via similarity analysis methods into super contexts S_c. The methods that might be applied for similarity reasoning can be statistical, probabilistic (e.g. Baysian), possibilistic (e.g fuzzy sets) or machine learning and data mining based (e.g. decision trees). Aggregation into super sets is done to *mitigate collection of missing or imperfect information and to minimize computational overhead when processing contexts.*

definition: A context is a collection of attributes aggregated into a feature vector describing a natural or abstract event.

A super context can be described as a triple denoted by:

$$S_n = (C_n, R_n, S_n)$$

where C is the context data of multiple feature vectors, R is the meta-data processing rules derived from the event and contexts data and S is controls security processing. S is defined to be a feature vector in this model that holds information about security levels elements or including overall security level requirements.

definition: A super context is a collection of contextual data with a feature vector describing the processing of the super context and a security vector that contains security level and other types of security information.

The cardinality of F with C is:

m:1

which when substituted into S creates a (C, R, S) cardinality of:

m:1:1

for the proposed model. However, we have not examined the impact, constraints of implications of having an

m:n:o

type of cardinality.

All of the above are a *type* of feature vectors where the elements of the vector can contain any type of information including the derived contextual processing rules and security methods for the given super context.

A super context is composed of context data from many sensing event objects, Eo_i,. As such contextual information collection works in a similar fashion to sensor networks and can borrow from theory in the field. Figure 1 shows the nature of collection of event object data over time. One can visualize a region of interest, e.g. the Indian Ocean tsunami for which event object data is collected which is centered over a thematic event object. In this case a thematic object when one considers all the data that may exist for the Indian ocean is the concept of the *origin of the tsunami*.

definition: A thematic event object (Teo) is the topic of interest for which event objects are collecting data. An example of a Teo would be the center of a tsunami.

In previous work [9], objects motions where characterized and described based on temporality, spatiality, impact and similarity. Development of these classes then lead to a grammar which derived rules that could have processing actions assigned to them. This allowed the notion of context to produce the paradigm of contextual processing. Simply put, *the nature of the information controlled the operation of the processing.*

3. SPICULE AUTENTICATION

The Spicule visual state change detection method[8] was originally conceived to be a simple and intuitive way to detect intrusions on computer systems. Years after its conception it was discovered that it had a variety of interesting applications based on the mathematics behind the paradigm. One of these turned out to be the ability to generate spatial authentication signature faster than standard hashing and encryption methods.

The Spicule's mathematics is based in vector algebra, and thus there is an algebra that exists for comparing two Spicule's to detect visually state changes in system state variables. Specifically, if the mathematical representation

of two Spicule's is subtracted a "change form" is created. The change form can be visualized which then results in a smooth featureless 3D ball if the two versions of the Spicule authentication signature are similar. The advantage of this is that it is simple and visually intuitive to recognize change with out having to conduct analysis or inspection of the underlying mathematica1 data. Figure 1 shows an example of the Spicule.

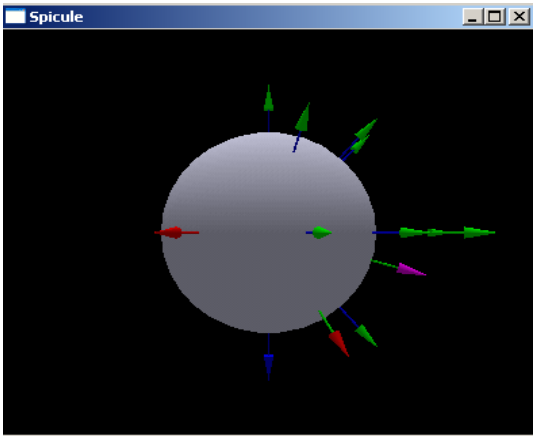

Figure 1. Sample picture of the Spicule

The development of the Spicule for authentication started with some research by Takeyama and Couclelis. They demonstrated that the GIS layering abstraction of a location is equivalent to a set of multiple attributes [9]. This means that various attributes about the same spatial object could be modeled that that selection of similar classes of attributes for a range of objects to be modeled for a variety of applications including authentication. This layering can be conceived as as a 3D-set of layers on top of each other.

In the layering paradigm, the Spicule can be utilized to create a mathematical *signature* for authenticating spatial data by mapping the tips of vectors on the Spicule to the unique spatial objects identified from the taxonomy. The signatures that can be generated using this approach becomes an n tuple which can be visually subtracted using Spicule to detect changes in the spatial data. This n tuple consists of information about a specific spatial objects vector consisting of a unique set of attributes such as magnitude, angular orientation and location of a vector on the 3D central ball. The vectors can be mapped to objects in various layers of spatial data objects (mentioned above), thus creating vectors that are not tied to the objects in a given layer, increasing the uniqueness of the signature. The number of vectors going from the Spicule was equal to the number of selected objects from the spatial dataset being authenticated. The collection of these vectors for a given set can be then used to describe a unique signature for a particular GIS data set.

The idea behind the developed authentication process was to utilize the Spicule tool to create a geometrical vector for each of several spatial objects selected from the spatial data sets. Vectors can be point from the center of the Spicule to the (x, y, z) coordinates of a spatial object. Each vector is thus unique and has three attributes that are represented as follows:

$$V_i = (\text{degreesVertical}, \text{degreesEquator}, \text{magnitude})$$

In this scheme there is a vector pointing from the center of the Spicule, at the origin, to each point or spatial object selected from the spatial data set for signature generation. In the previous work three data layers are initially proposed to be placed at one vertical unit apart from the Spicule layer. So, the first layer points will have coordinates of (x, y, 1), the second layer points' coordinates will be (x, y, 2), and the third layer points' coordinates will be (x, y, 3). Based on this the vector attributes for each authentication point in the three layers were:

$$Mag_i = \sqrt{x^2 + y^2 + i^2}$$

where:

i is the data layer number.

x, y are point original coordinates.

Mag_i is the magnitude of the vector from (0,0,0) to a point in layer i.

$$Sin\theta_{ei} = \frac{x}{\sqrt{x^2 + y^2}} \Rightarrow \theta_{ei} = Sin^{-1}\frac{x}{\sqrt{x^2 + y^2}} \qquad (2)$$

$$Sin\theta_{vi} = \frac{i}{\sqrt{i^2 + y^2}} \Rightarrow \theta_{vi} = Sin^{-1}\frac{i}{\sqrt{i^2 + y^2}} \qquad (3)$$

Equations (2) and (3) are used to calculate the equator and the vertical angles respectively,

where:

i is the data layer number.

θ_{vi} is the vertical angle degrees for a vector from (0,0,0) to a point in layer i.

θ_{ei} is the equator angle degrees for a vector from (0,0,0) to a point in layer i.

The collection of attributes and angles for all authentication vectors forms a two-dimensional matrix that is used as for the authentication signature and the Spicule visualization authentication process.

The signature calculation process is done when a spatial dataset is requested to be transmitted over the internet. Table 3 shows a sample calculated vector matrix.

Object ID	Layer	Mag_i	θ_{vi}	θ_{ei}
1	3	7.68	66.8	18.43
2	2	16.31	42.51	4.76
		. .	. .	. .
n	i	29.22	51.95	3.18

Table 3. Sample calculated vector matrix

At the receiving end, the same process to create a signature matrix from the *received* spatial dataset was applied. By visualizing the mathematical difference between the received spatial data sets matrix and the transmitted matrix, it can be determined if the dataset has been intercepted or altered during transmission. This process may be described by:

IF Visual Mathematical Difference = 0 THEN

No Interception or Alternation.

This is the standard logic found in traditional authentication schemes. In the above method if the visual mathematical *difference* (vector based subtraction) between the two matrices does not equal to zero, it is assumed that the spatial dataset has been intercepted and altered. However, we can not determine the extent and the type of change that have been made because removal, addition, or movement of a given spatial object or point may result in the change of sequence for many vectors in the matrix after the point of modification in the matrix. *The nice thing about application of Spicule is that visualization of the signature matrices with Spicule and application of visual subtraction of the vectors results in a Spicule devoid of vectors if the data objects have not been moved or modified during transmission. The intuitive nature of this visualization makes it easy for an analyst with the most basic of skills to determine if data has been modified and how much.*

4.0 COMPARATIVE AUTHENTICATION SIGNATURE GENERATION PERFORMANCE

Spatial data may be protected for transmission by encryption or by the generation of a signature using MD5, SHA or RIPEMD. In order to compare the performance of the spatial signature approach to that of above traditional methods a test suite was set up on a PC running at 2.4ghz with a P4 processor. The Crypto++ package was utilized for comparison with timing figures measured down to the millisecond. Crypto++ has a program call Cryptest that may be called with command line switch to encrypt symmetrically, decrypt and generate SHA, MD5 and RIPEMD160 digests. The command line interface was invoked from a command line shell generated with Visual Studio. Because Cryptest was being called using a system command from inside the compiled test program, the first part of the test suite called the operating system shell to load a simple C program. This allowed us to measure the effect on performance of just loading a simple program. Of note in the spatial signature generation test, this test selects increasingly more and more static spatial objects from the test data which are part of the objects from the previous work with taxonomies mentioned above. The above test was run thirty times for each part of the above test program with the following results:

Test Type	Pass 1 (10x)	Pass 2 (10x)	Pass 3 (10x)
Shell	63.00	58.00	57.00
Encrypt (symmetric)	126.60	123.4	121.90
Decrypt (symmetric)	115.60	123.5	121.90
MD5/SHA/RIPEMD	67.20	67.20	64.00
Spatial Authentication	< .01 millisecond	< .01 millisecond	< .01 millisecond

Table 4 Average performance comparison

of Spatial Authentication versus Symmetric encryption, SHA, MD5, RIPED (milli seconds) on test data

4. SPATIAL AUTOCORRELATION APPROACH

Spatial autocorrelation was developed by Moran in 1995 and has the potential to integrate contextual modeling in such that a reduced number of spatial objects can be selected for the Spicule authentication. This section discusses how such a method may work and is the subject of future research.

Global spatial autocorrelation measures the degree to which objects on a spatial grid are related to other objects. The notion is based on spatial dependence which can be defined as "the propensity of a variable to exhibit similar values as a function of the distance between the spatial locations at which it is measured"[7]. Put more simply, the value of a spatial variable is often influenced by its neighbors.

Global spatial autocorrelation can be defined given variable $x = \{x1, \ldots, x_n\}$ sampled over n locations[7]. Morans spatial correlation coefficient can be calculated by:

$$I=\frac{zWz^t}{zz^t}$$

where:

$z = \{x_1 - \bar{x},\ldots x_n-\bar{x}\}$

z^t – is the transpose of z

W - is a rectangular row normalized contiguity matrix

Localized spatial auto-correlation (LSA) is similar to global autocorrelation. Instead of measuring the correlation of a group of objects at a global level, it is a measure that determines how correlated a given variables location might be correlated and influenced by its neighbors. This is a derivation of I and is given by:

$$I_i=\frac{z_i}{s^2}\sum_j \frac{W_{ij}}{z_j},\ i \neq j\ ,$$

where:

$z_i = x_i - \bar{x}$

s – is the standard deviation of x

W_{ij} - is the contiguity matrix, normalized, or based on similarity

The application of local autocorrelation and context might follow the logic that

i) a user wants to retrieve object for a given location in space and or in a given time period for that location.

ii) the object the user might want to look at are of a given class with heterogeneous members. For example:

O = {tank, half trac, jeep, jeep with gun mount, armored personal carrier}

where:

O – is object class of battlefield objects with wheels

Note that within this class there are implications for *similarity* from the context model such as *members that can fire projectiles* and *members that transport resources*.

These members will have spatial locations, temporal loci and impact relationships for a given location, T_{eo} and for other given themes that might be part of a retrieval query such as *fighting, moving, transporting*.

To demonstrate how LSA might be integrated with context, consider the follow example. Figure 2 is spatial lattice where members of **O** are located in various concentrations and dispersions.

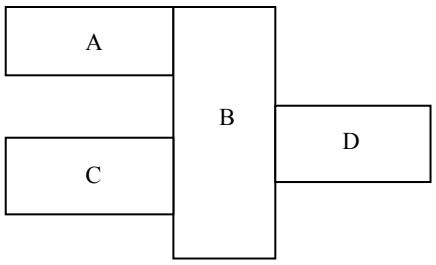

Figure 2, A contiguity lattice C of associated cells over a spatial extent with members of set O dispersed in various cells of the lattice.

Considering the argument of spatial dependency, one can see that concentrations of vehicles with guns can tend to be related among adjacent cells in the lattice, and that the same could apply for concentrations of vehicles that are used for transport.

The LSA method takes the above lattice and constructs a contiguity matrix as shown in figure 3. This matrix is the beginning of the LSA and identifies which lattice cells have shared edges and thus my have correlations among the cell contents.

	A	B	C	D
A	0	1	0	0
B	1	0	1	1
C	0	1	0	0
D	0	1	0	0

Figure 3, Contiguity Matrix M_c

Use of variations of the contiguity matrix in the Spicule approach is going to be the subject of further research and development, however, the current LSA uses a normalized matrix such as shown in Figure 4. Normalization is down to minimize the undue influence on calculation of I_i due to a large number of contiguous cells around a cell of interest.

	A	B	C	D
A	0	1	0	0
B	.3	0	.3	.3
C	0	1	0	0
D	0	1	0	0

Figure 4, Row Normalized *W*

Application of the LSA I_i can now provide the basis for application to the Spicule authentication method.

Consider that a user is interested in query Q_1:

Q_1 = *(the location of the majority vehicles with guns on them*, T_{eo})

Q_1 is a very realistic type of query for planning attacks or logistics. The steps to apply LSA in this type of query would be:

i) build *C*

ii) build $M_c = fn(C)$

iii) calculate $W = fn(M_C)$

iv) calculate $I_i = fn(W)$

v) apply Q_1 for some sort of selection criteria producing **O**

vi) generate authentication signature vector

$$s[] = Spicule(O)$$

Application of the above if done properly could produce a reduced number of spatial objects to authenticate and thus improve the already fast processing of Spicule.

Step v in the above algorithm implies some sort of selection method on the correlation coefficients I_i . This can be done by application of one of the following criteria:

- similar values,
- above a *floor* value,
- below a *ceiling* value
- falling into a bounded range

As an example, consider calculated lattice consisting of localized correlation coefficients for Q_1 as shown in Figure 5, a selection criteria for correlation of .8 ± .2, and a region of interest T_{eo}. Calculation of a localized correlation values might result in the following type of lattice where

	A	B	C	D
A	0	.82	0	0
B	.79	.8 T_{eo}	.5	1
C	-.2	.23	.4	0
D	0	1	-.6	0

Figure 5 Sample calculated local correlations I_i over *W*

the T_{eo} is a spatial location that Q_1 is centered upon. The application of the algorithm presented would result in the selection of lattice cells containing {.82, .79 and .8}.

A research note for the future would be to examine performance of this method and how it degrades as the granularity of the lattice increases. The current method proposes lattice cells that contain sets of objects, however it would be possible to say that each cell is a single object.

This LSA approach has some very handy properties when considering the CP model and integration into the Spicule method. The first of these is that it

i) integrates the dimension of spatiality into the Spicule process. It does this by organizing objects into lattices and incorporating the notion of spatial dependency. It also does not force unnecessary overhead on the calculation of I_i because it operates based on the notion of an irregular lattice, not a fixed lattice or a fixed grid. Cells without objects meeting Q_1 are merely ignored in the method.

ii) LSA has the built in notion of spatial dependency. This again is the idea that what is close to you spatially probably affects the value of adjacent cells. In the above example we have argued that it could be the case that vehicles with guns may be concentrated in certain lattice cells. This allows the proposed approach to integrate nicely with the CP dimension of *Similarity* . This idea might be defined as

Similarity based on spatial dependency

8. FUTURE WORK

The presented method integrates the new CP model with Spicule authentication via application of the LSA approach to create a new contextually based authentication paradigm. At present much work is being done on the theoretical constraints and applications of these methods. This leads to the opportunities for much more empirical work to be considered. Future research issues are many. One may investigate:

i) the integration of impact and time into the above proposed method and how that may be modeled.

ii) how granularity of lattice cells affect performance in selection of objects for Spicule authentication

iii) Boolean algebras for combinations of selection criteria in the proposed algorithm in this paper.

iv) $S_n = (C_n, R_n, S_n)$ and what the security term S_n may be defined as based on the method proposed in this paper.

CP has proven to be a useful paradigm in several areas of computer science and should continue to be investigated to develop a significant corpus of knowledge.

8. REFERENCES

1. Rosemann, M., & Recker, J. (2006). "Context-aware process design: Exploring the extrinsic drivers for process flexibility". T. Latour & M. Petit *18th international conference on advanced information systems* engineering. proceedings of workshops and doctoral consortium*: 149-158, Luxembourg: Namur University Press.*
2. Schilit, B.N. Adams, and R. Want. (1994). "Context-aware computing applications" (PDF). *IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'94), Santa Cruz, CA, US*: 89-101.
3. Schilit, B.N. and Theimer, M.M. (1994). "Disseminating Active Map Information to Mobile Hosts". *IEEE Network* **8** (5): 22–32. doi:10.1109/65.313011.
4. Dey,Anind K. (2001). "Understanding and Using Context".*Personal Ubiquitous Computing* **5** (1): 4–7. doi:10.1007/s007790170019.
5. Cristiana Bolchini and Carlo A. Curino and Elisa Quintarelli and Fabio A. Schreiber and Letizia Tanca (2007). "A data-oriented survey of context models" (PDF). *SIGMOD Rec.* (ACM) **36** (4): 19--26. doi:10.1145/1361348.1361353. ISSN 0163-5808. http://carlo.curino.us/documents/curino-context2007-survey.pdf.
6. Schmidt, A.; Aidoo, K.A.; Takaluoma, A.; Tuomela, U.; Van Laerhoven, K; Van de Velde W. (1999). "Advanced Interaction in Context" (PDF). *1th International Symposium on Handheld and Ubiquitous Computing (HUC99), Springer LNCS, Vol. 1707*: 89-101.
7. Shekhar, S; Chawla, S;(2003). *Spatial Databases A Tour*, Prentice Hall, p 190.
8. Vert, G.; Iyengar, S.S.; Phoha, V.;(2009) *Security Models for Contextual Based Global Processing an Architecture and Overview*, Cyber Security and Information Intelligence Research Workshop, published in ACM Digital Library, Oakridge National Laboratory, TN,.
9. Vert G; Phoha, V; Iyengar, S.S; (2010). *Contextual Processing Theory and Applications*, Taylor and Francis.