

Latent Knowledge Structures of Traversal Behavior in Hypertext Environment

PERWAIZ B. ISMAILI

School of Behavioral and Brain Sciences

University of Texas at Dallas

UNITED STATES OF AMERICA

pbi021000@utdallas.edu

Abstract: - In this paper, we introduce “Knowledge Diagram Contribution” (KDC) analysis as a novel categorical time-series method in observing underlying traversal knowledge structure of experts by exploiting varying hypertext (web) presentation formats and knowledge domains. The navigation behaviors were studied by designing hypertext presentation formats and domain text that adheres to content design principles inspired by discourse and text comprehension scholars. As a continuation of previous study by Ismaili & Golden [1], twenty undergraduate psychology students from University of Texas at Dallas participated in this study. Students traversed through different Hypertext (web) presentation formats while reading content from three different knowledge domains controlled for micro (web-page, web-site) and macro (consistent semantic connections across knowledge domains) characteristics. The influence of expertise and web traversal behavior in deriving underlying knowledge structures is presented using KDC analysis. In addition, previously reported Classical data analysis (ANOVA) are compared with KDC analysis in highlighting quantitative and qualitative differences of these derived latent knowledge structures. As compared with novice, experts tend to exhibit sequential and semantic traversal patterns across all three web formats, whereas, novices are more influenced by and therefore tend to employ random navigation strategies.

Key-Words: - Hypertext, Knowledge Structures, Navigation, Traversal Patterns, Content Design, Expertise

1 Introduction

Reading is fundamental, as commonly reported, and is a complex process that requires several crucial interim steps before a reader extracts meaning from the presented text. Therefore, this series of transformation from surface text to a higher situational level [2] or from textual subsystems to larger meta-system [3] is extremely important for coherent understanding of the text. An essential component of successful reading is to create a multilevel representation of texts i.e., processing from low level (individual words) to high level where the gist of the presented information are derived. These processes have to work in concert before the meaning from the texts are extracted. Kintsch (1988) identified three levels of text representation: the surface level, the textbase level and the situation model. Therefore, it requires the readers to decode words, integrated individual word and sentence meanings into a coherent representation of text. In other words, the involved process of reading at a surface level extracts words and syntax which are formalized as propositions, to preserve its meaning at text base level which is then transformed into coherent understanding of text at the situational level which represents the global

meaning of text incorporating reader’s prior knowledge. In addition, during the reading process, not all the information that is needed to comprehend a passage is presented in the text by the author. Moreover, the inclusion of all such details would greatly obstruct and obfuscate the reading and comprehension processes. Therefore, it is considered beneficial to omit these superfluous details. Research from Discourse and Text comprehension suggest that readers depend on hierarchy [4] or structural patterns at the local and global level in order to recognize the type of text and integrating relevant parts for better comprehension [5], and construct a coherent understanding of a text [6]. Inferences make assumptions about the reader’s internal representation of the text and can be represented as structural patterns. Hence these patterns may be modeled as semantic networks or knowledge structures, where concepts are represented as nodes and relationships between these concepts are represented as connections [7]. Furthermore, various research suggest that compared with subordinate concepts, super-ordinate nodes [8] are likely to be recalled and nodes with more semantic connections [9] are more memorable. For example,

Trabasso et al (1984) shown just that, high density connections compared with fewer interconnections between nodes are perceived comparatively more coherent. Moreover, experimental evidence from discourse literature suggest the importance of the order in which nodes are stated in inference production data which is deemed to be revealing of the structural characteristics of the reader's mental model, therefore, essential and sensitive to readers underlying mental representation of text [for example, 9, 24 & 25].

Although research in text comprehension has focused on sequential, sentence-by-sentence or paragraph-by-paragraph texts which can be categorized by linear, print-like formats, however, shed valuable insights in further exploring the alternate web formats used for information presentation. More recently, important alternative versions of texts such as hypertext have become widely used as a result of the wide-spread use of the World Wide Web as a means of communicating complex information. The authors in these multidimensional web space are challenged in adequately guiding its reader through wealth of diverse type of information, sources and services, while minimizing the effects of feeling confused or lost commonly attributed to 'cognitive overload'[10]. Furthermore, experience with the technology, presentation formats and navigation through non-sequential websites causes many reading comprehension situations especially for novices unfamiliar with the knowledge domain.

The introduction of WWW, hypertext and other electronic medium opened door for the widespread use of hypermedia technology in the education which, by nature thrives on a rich non-linear design where multiple concepts and resources are easily accessible and allows tapping into wealth of knowledge, potentially deemed critical for learning. Much of the excitement was about the wealth and type of the information repositories of interconnecting, interactive knowledge [32], the real life-like [33] nature which mirrors the organization of how people organize concepts of the world, and, was thought to have the potential of changing the nature of reading. All that however, at a cost of comprehension which supposedly should have been a major performance indicator of learning experience using technology [34], whether the information is structured linearly or non-sequentially.

Past research suggest that minimal non-linear formats are expected to help readers with latent concepts [12], therefore, it needs to be exploited. However, research have demonstrated that non-

linear, hypertext Web presentations impacts learning performances negatively compared to traditional, print-like linear Web designs [11]. Recent evidence suggests that nonlinear designs may facilitate learning of the interconnections (structure) of the presented information [13].

Different navigation patterns have been noticed with varying levels of reader knowledge [14]. In addition, domain knowledge plays a significant role in predicting the recall and improved comprehension for certain type of readers [15]. However, results are not conclusive and sometimes contradictory. For example, some research show positive effects of prior knowledge [16], whereas other research has shown no effects of prior knowledge in assisting comprehension for low knowledge readers in hypertext environments. For low knowledge readers, the hypertext environment may sometimes be perceived as "confusing" and a linear or structured semantic based may be more appropriate for supporting learning and comprehension. This latter goal might be achieved by designing hypertexts which more effectively exploit the semantic organization (interconnections) which the author wishes to communicate to the reader.

Classical analyses are prone to neglect lots of information presented in the observed data, mainly the order in which readers may have traversed in a web environment. The underlying cognitive process of behind comprehension of an expert can be inferred by the order in which web pages were visited [28]. KDC analysis is a new analysis tool that not only takes into account the order in which nodes are visited; it also simultaneously calculates the probability of the ordered sequence. KDC analysis uses these probabilities to determine the support provide by the traversal data for each of the proposed theories (knowledge structures or digraphs) it is considering during analysis.

Golden [18], introduced a particular type of data analysis technique called Knowledge Digraph Contribution (KDC) analysis which exploits specific types of semantic networks called "knowledge digraphs" (or Knowledge Structures). KDC analysis is based upon the assumption that the connections in a knowledge digraph can be thought of as a path the inference process is likely to follow and this inference process influences the order in which concepts or ideas are processed during comprehension, recall or web traversal. In this paper, KDC analysis is used in order to study the effects of traversal patterns in different hypertext presentation formats.

In general, classical sequential data analysis methods are not typically used by discourse and text comprehension researchers to analyze sequential statistical regularities in identifying web site traversal patterns of readers. Moreover, even if such sequential techniques were used more extensively [19] it would not be sufficient. Since these methods tend to suffer from over-fitting problems due to its inability to incorporate constraints upon patterns of associative strengths within these types of models such as semantic based networks .

In order to address these issues by accommodating confirmatory analysis, Golden [18] developed a dynamic constrained parametric multinomial time-series regression model for categorical time-series analysis for any ordered sequential data analysis. Knowledge Digraph Contribution (KDC) analysis models allows researcher to specify a collection of models (digraphs or structures) representing different types of semantic relations, constraints and networks. Given such a set of “knowledge digraphs”, KDC analysis estimates (using maximum likelihood estimation) a “contribution weight parameter” for each directed graph (digraph) for the purposes of predicting/explaining orderings of such sequential ordering of web pages traversal. Additionally, KDC analysis has a distinct advantage over classical sequential data analysis methods because all of the asymptotic statistical tests developed using KDC analysis are derived within model misspecification which permits reliable statistical inferences even when the theoretical assumptions about the types of semantic relations among models are not entirely correct [20].

KDC analysis may be used to refine and develop theories of directional semantic connectivity as represented by digraphs. For example, while exploring and comparing the knowledge structures theories within text comprehension domain, Golden showed that the forward link causal model fit the free recall data significantly better than a causal model incorporating both forward and backward links. This example suggests that standard statistical methods of free recall data analysis are sometimes not sufficiently sensitive in capturing crucial statistical regularities since they do not explicitly incorporate the theorist’s conception and directionality of the critical semantic connectivity patterns within the nodes of the network.

2 Specific Aims

Human society is now heavily dependent on the world-wide web as a communication, learning, commercial medium. Therefore, it is critical to explore this area for variety of purposes such as, e-commerce [26] and distance learning [27] to name a few. The “web” being a complex, multidimensional and dynamic multimedia environment requires systematic scientific approach to unravel its underlying potentials as a optimal medium for services such as online learning. Following the recent work on comprehension in hypermedia environment, discourse and to further continuous exploration by Ismaili & Golden [1], this study investigated whether latent knowledge structures be derived from traversal behavior including its influence when reading takes place in two different albeit qualitatively similar types of hypertext environments (semantically organized versus fully connected) relative to a linear text environment. Furthermore, this research investigated the influence of domain knowledge expertise by considering twenty undergraduate Psychology students for this study. These students navigated web sites which are in the knowledge domains of Psychology, Neuroscience, and (as a control condition) Archeoastronomy. Traditional (analysis-of-variance) data analysis methodologies was used to establish the presence or absence of phenomenon, compared with highly sophisticated KDC analysis in deciphering latent traversal structures.

This study being the extension of previously reported research [1], therefore, in this article some of the pertinent information is briefly repeated in the next section as a general background for better understanding and appreciation of results presented in this article using sophisticated KDC analysis.

3 Background

3.1 Content

As reported by Ismaili & Golden (2008), paragraphs of three science texts, referred to as three knowledge domains, (Psychology, Neuroscience and Archeoastronomy as a Neutral text) were designed with each paragraph containing two sentences. Each paragraph represents a web-page of a particular knowledge domain web-site. The semantic associative relationships among the topic sentences of each of these paragraphs had qualitatively same patterns across all three knowledge domains. Please see Figure 1 for the content design template and Table 1 for sample of Psychology web page. Furthermore, at micro-level, meticulous attention

was given in order to ensure the balanced content design such as: word count, level of difficulty, average number of words per web-page and Average number of words per web-site. The idea was to keep stimuli characteristics across all three knowledge areas (Psychology, Neuroscience and Archeoastronomy) qualitatively equivalent at both micro and at macro web-site level. Each of knowledge domain consisted of five facts, two intermediate conclusions, two irrelevant fact nodes and a final conclusion. The two Irrelevant Fact nodes were mutually associated however not with the main semantic content chain. (see Figure 1).

In short, from Figure 1, the connections between web-pages (nodes) depict semantic (logical associations) linking between nodes. For example, consider following connections:

F1 → F3 and, F2 → F3.

It suggests that Fact node-3 is logically connected with the topic sentence of Fact node-1 and with the topic sentence of Fact node-2. See Ismaili & Golden [1] for more details.

Figure 1: All three knowledge domain (Psychology, Neuroscience & Archeoastronomy) area content were kept consistent for semantic associations between nodes and in consisting five ‘fact’ nodes (F1 – F5), two ‘irrelevant fact’ (IF-1 & IF-2) nodes, and, three nodes presenting two ‘intermediate conclusions’ (IC-1 & IC-2) and a ‘final conclusion’ (FC).

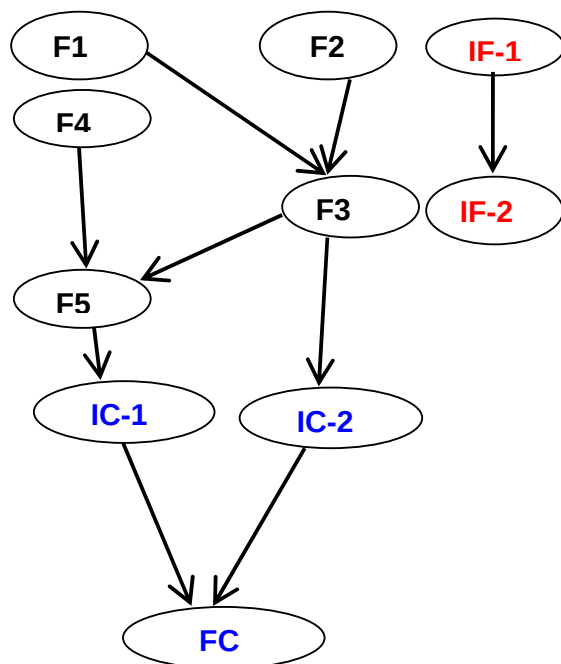


Table 1: Sample Psychology text read by participants. All three knowledge domain (Neuroscience & Archeoastronomy) web sites consist of ten web-pages: Five web pages for Facts 1-5, two web-pages for the two

Irrelevant Fact nodes, and, three web-pages presenting two Intermediate conclusions and a Final conclusion. Only one topic sentence has been shown. For more see Ismaili & Golden [1].

Psychology Content

F1 (Fact 1)- Craik and Lockhart proposed that stimuli subjected to semantically processed stimuli (“deeply processed stimuli”) will be more memorable than perceptually processed stimuli (“shallow processed stimuli”)...

F2 (Fact 2)- Tulving proposed that memory performance increases as the similarity between encoding and retrieval contexts increases...

F3 (Fact 3)- Although many scientific studies have reported experimental results supporting both LOP and ESP, some research has identified situations where LOP theory fails while ESP theory is successful...

F4 (Fact 4)- Although these early context-independent memory models were highly influential, later experimental findings showed these early models could not account for experimental findings as effectively as context-dependent models such as LOP and ESP, since they ignored encoding and retrieval factors...

F5 (Fact 5)- The psychologist Abernathy (1940) studied context effects on test performance and showed memory recall performance improved when students were tested in the same physical environment as the environment in which they received test instructions...

IF1 (Irrelevant Fact 1)- One traditional theory of long-term memory storage assumes that items that are retained in short-term memory for extended periods of time will eventually be transferred to long-term memory...

IF2 (Irrelevant Fact 2)- Associative theories of long-term memory storage postulate that items with more semantic connections are more effectively retained...

IC1(Intermediate Conclusion 1)- ...Accordingly, ESP models are considered preferable over LOP models.

IC2(Intermediate Conclusion 2)- ...It is now widely recognized that memory models need to incorporate context-dependent factors.

FC(Final Conclusion)- In summary, ESP models are preferable to LOP models because they incorporate interactions between encoding and retrieval factors. Therefore, ESP models are context-dependent memory models which emphasize that memory performance is specific to the environment in which it is embedded.

3.2 Hypertext (web) Design

Towards this end, three web site formats: Linear, Meshed-hypertext, Semantic-based were created for each of the three knowledge domains (Psychology, Neuroscience, and Neutral text-Archaeoastronomy). A navigation bar was provided for all of these three web formats on the left hand side of the web pages with buttons to go directly to the introduction page, the end page, and all the interconnected nodes, with some variations.

Although outside the scope of this paper, the linear navigation format encouraged sequential traversal while in the Meshed environment (Figure 3), subjects were free to move around between pages using navigation bar as they wished. In the semantic-based web site format subjects were allowed to move among all pages, but the navigation scheme included suggested traversals as solid lines between nodes (web pages) depicting recommended traversal paths (since they showed semantically associated nodes). The semantic-based hypertext presentation was created using the outcome of the content design as depicted in Figure 1. Although, participants read three different text (psychology, neuroscience and neutral-archaeoastronomy), only two experimental texts were analyzed for this article.

3.3 Knowledge Digraph Contribution-KDC

As previously noted, KDC analysis is based upon the assumption that readers (experts and novice) are likely to follow the certain paths through the hypertext (web) environment. Therefore the directional flow of the knowledge digraph (knowledge structures) representation should be predictive of sequence of pages they are likely to follow. Although, there are a variety of ways in which navigational analyses can be performed, there are two ways in particular which are consistent with the existing experimental literature. The first method involves analyzing or counting the number of times each page was visited, (i.e., Presence of absence of effect) which, is relatively straightforward using classical ANOVA analyses. The other method involves analyzing knowledge digraph (knowledge structures) representations by looking at the order in which nodes were traversed sequentially in the web environment.

Assessment for sequential order of web page traversal in KDC was performed by creating three different knowledge structures (digraphs) as an ordered pair of nodes in the knowledge digraph (specified by a directed link or arrow), to represent expected linear, random and semantically ordered navigation patterns.

Golden [18], introduced a particular type of data

analysis technique called Knowledge Digraph Contribution (KDC) analysis which exploits specific types of semantic networks called “knowledge digraphs” (Knowledge Structures). KDC analysis is based upon the assumption that the connections in a knowledge digraph can be thought of as a path the inference process is likely to follow and this inference process influences the order in which concepts or ideas are processed during comprehension, traversal or production which could be bidirectional. In this research, KDC analysis is used in order to study the effects of traversal patterns in different presentation format in the hypertext environment. In addition it features creation of digraph as realistic models of latent knowledge structures for further verification, and confirmation. Furthermore, KDC methodology minimizes over-fitting problems as observed in classical sequential data analysis. Although classical sequential data analysis are not typically used for sequential web navigation pattern analysis, but even if these techniques [30, 31] were used it would not have appropriate for this type of investigation. Classical sequential methods do not support mechanisms for estimating which knowledge digraphs, out of three (linear, random & semantic) best explains the traversed website patterns.

Golden [18] developed a highly constrained parametric multinomial time-series regression model for categorical time-series analysis of free response data as an ordered sequence of propositions which are applied for this study towards studying the different traversal patterns. Golden (1998) refers to models of this type as Knowledge Digraph Contribution (KDC) analysis models allows representation of different types of semantic relations among propositions. Given such a set of “knowledge digraphs”, KDC analysis estimates (using maximum likelihood estimation) a “contribution weight parameter” for each directed graph (digraph) for the purposes of predicting/explaining orderings of propositions. Additionally, KDC analysis has a distinct advantage over classical sequential data analysis methods because all of the asymptotic statistical tests developed using KDC analysis are derived within the general theory of model misspecification which permits reliable statistical inferences even when the theoretical assumptions about the types of semantic relations among for example nodes (or propositions) are not exactly correct.

More formally, assume the i th participant in the study generates a finite sequence of T_i propositions (or nodes) represented as the ordered sequence of d -

dimensional vectors: $\mathbf{f}_{i,1}, \mathbf{f}_{i,2}, \mathbf{f}_{i,3}, \dots, \mathbf{f}_{i,T_i}$. If the m th proposition in the proposition dictionary is the t th proposition mentioned by participant i , then $\mathbf{f}_{i,t}$ will be the m th column of a d -dimensional identity matrix. The columns of the d -dimensional identity matrix are sometimes referred to as the *proposition* (or *node*) *dictionary*. The sequence $\mathbf{f}_{i,1}, \mathbf{f}_{i,2}, \mathbf{f}_{i,3}, \dots, \mathbf{f}_{i,T_i}$ is a particular realization of a “ τ -dependent” stationary stochastic process: $\tilde{\mathbf{f}}_{i,1}, \tilde{\mathbf{f}}_{i,2}, \tilde{\mathbf{f}}_{i,3}, \dots, \tilde{\mathbf{f}}_{i,T_i}$. The d -dimensional square matrix $\mathbf{D}_q^{(k)}$ is used to denote the k th digraph with time-delay q . For example, a causal digraph might be represented using this notation by a matrix $\mathbf{D}_q^{(k)}$ such that the ij th element of $\mathbf{D}_q^{(k)}$ is equal to one if a causal link specifies that the j th proposition in the proposition dictionary is the causal antecedent of the i th proposition in the proposition dictionary for a particular text (the index k would identify the semantic label for the digraph which in this case is “causal digraph”). The contribution weight associated with digraphs $\mathbf{D}_1^{(k)}, \dots, \mathbf{D}_L^{(k)}$ will be denoted by the real scalar parameter $\beta^{(k)}$. The parameter vector for the KDC probability model is thus denoted by a M -dimensional real vector $\boldsymbol{\beta} = [\beta^{(1)}, \dots, \beta^{(M)}]$. The positive integer L is referred to as the KDC model’s *working memory span parameter*.

The KDC Markov model is specified within a Bayesian framework by the following constrained multinomial logistic regression time-series model:

$$p(\mathbf{f}_{i,t} | \mathbf{f}_{i,t-1}, \dots, \mathbf{f}_{i,t-L}; \boldsymbol{\beta}) = \frac{\exp(h_{i,t})}{\sum_{j=1}^d \exp(h_{j,t})}, \quad h_{i,t} = \sum_{k=1}^M \beta^{(k)} \left[\sum_{q=1}^L \mathbf{D}_q^{(k)} \mathbf{f}_{i,t-q} \right] \quad (1)$$

with the prior on $\tilde{\boldsymbol{\beta}}$ is assumed to have a multivariate Gaussian prior with known constant mean vector $\boldsymbol{\beta}_0$ and known constant positive definite real symmetric covariance matrix $\mathbf{C}_\beta$.

Golden has shown that standard statistical methods of free recall data analysis are sometimes not sufficiently sensitive in capturing crucial statistical regularities since they do not explicitly incorporate the theorist’s conception of the critical semantic connectivity patterns among the propositions in the text (e.g., a causal network analysis of the propositions in the text). The same network connectivity of proposition has been applied in studying traversal patterns in the

Hypermedia environment by creating three KDC knowledge digraphs, namely: Linear (for sequential traversal patterns), Mesh (random patterns) and Semantic to represent the navigation patterns in a network of semantically connected nodes (web pages).

4 Methods

Twenty Psychology undergraduate students participated in this exploratory study. Each student acted as an Expert (reading Psychology text), as a Novice (reading Neuroscience text) and, as a control group participant by reading neutral Araeoastronomy text.

Each subject read all three knowledge domains i.e., Psychology, Neuroscience and Araeoastronomy as well as were exposed to all three Hypertext (web) presentation (Linear, Meshed and Semantic-based) formats. The order of web presentation format were counterbalanced across all the participants.

Each subject started with a Meshed Hypertext filler text – Astrophysics, which was not analyzed followed by the three counterbalanced experimental texts. After reading each website within allotted time participants were asked to summarize their understanding of the presented text before moving on to the next website. Each participants started with the ‘Introduction’ page and ended the website traversal by selecting ‘I am Done’ button in the bottom. Only participant traversal behavior analysis using classical ANOVA and KDC has been presented in this paper.

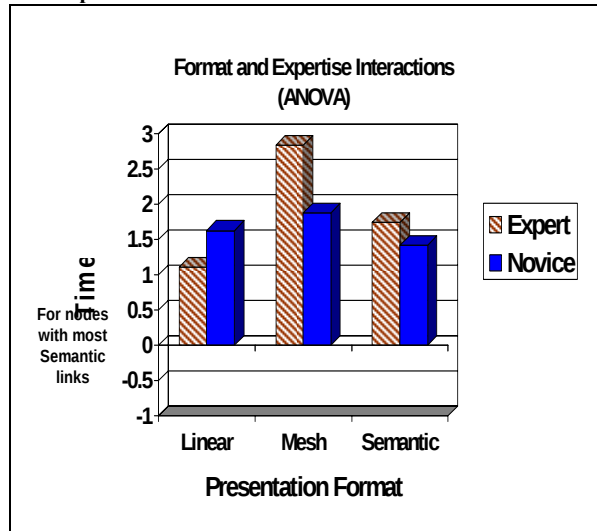
5. Results and Discussion

As previously reported by Ismaili & Golden [1, 21], two different types of performance data was submitted to an ANOVA with Expertise (Expert, Novice) as Between-Group, and Format (Linear, Meshed-Hypertext and Semantic-based Hypertext) as Between-Group variables. Only two experimental text (Psychology & Neuroscience) were analyzed for this study.

It was found that the total time spent on nodes with most semantic connections (i.e., F3 & FC) called ‘Link Nodes’ for presentation format F (2, 4) = 10.27, MSe = 4.044, $p=0.0002$, as well as for expertise interaction, $F(2, 4) = 2.90$, MSe = 1.142, $p= 0.0307$ reached significance. In general participants with strong expertise in knowledge domain were spending more time reading nodes with more semantic connections (link nodes) than other nodes especially in the Mesh-Hypertext

environment. Furthermore, experts were found to be spending more time visiting nodes with more connections as compared to novices in all three Hypertext (web) format ($p < .001$).

Figure 2: ANOVA analysis suggest that novice as compared with experts spent less time reading nodes with most semantic connections (link nodes) except linear format. Expert spent most time in Mesh-Hypertext and least in Linear presentation format.



ANOVA analysis although useful in highlighting the presence or absence of effect with statistical validity, however doesn't explore additional insights that may prove useful and be of importance to scientists in exploring and deciphering underlying performance structures such as traversal behavior of experts in the hypermedia environment [28]. For example, what navigation strategies were used by readers when presented with different web formats with varying knowledge expertise.

For the purposes of this preliminary data analysis report and in order to explore the underlying traversal structures of expertise and formats, we created KDC models which consists of three knowledge digraphs or structures for sequential navigation patterns (linear) between web-pages, a semantic-based digraph to explore traversal patterns of nodes that were associated semantically and finally a model representing random transitions referred to as a mesh-digraph. Please refer figure 3, 4, and 5 for KDC digraph analysis.

Figure 3: KDC Linear Digraph analysis show that when participants were presented with linear-hypertext (web) format, they regardless of expertise traversed sequentially. In addition, experts traversal behavior suggest semantic transition pattern between web-pages whereas novice behavior can be modeled using random Mesh-digraph.

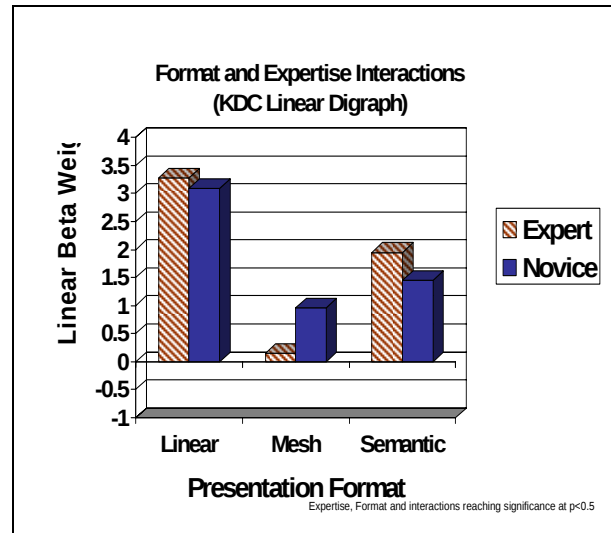


Figure 4: KDC Mesh Digraph analysis suggest that when readers were presented with mesh-hypertext format both novice and expert transition between nodes at random. However, experts seemed to be using ordered sequential navigation strategy between web-pages. Novice on the other hand transition between nodes randomly when presented with mesh format.

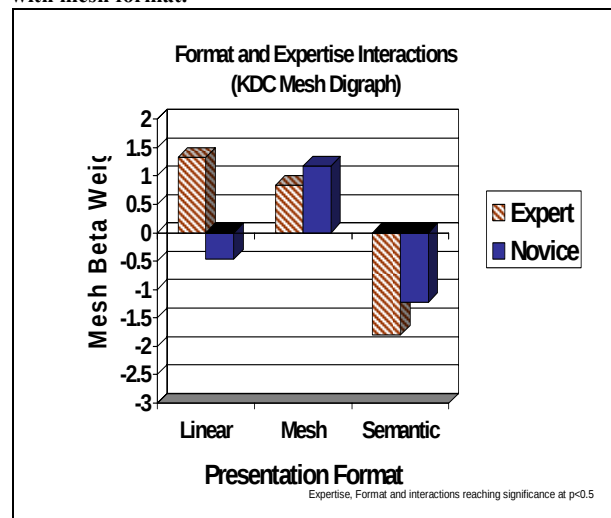
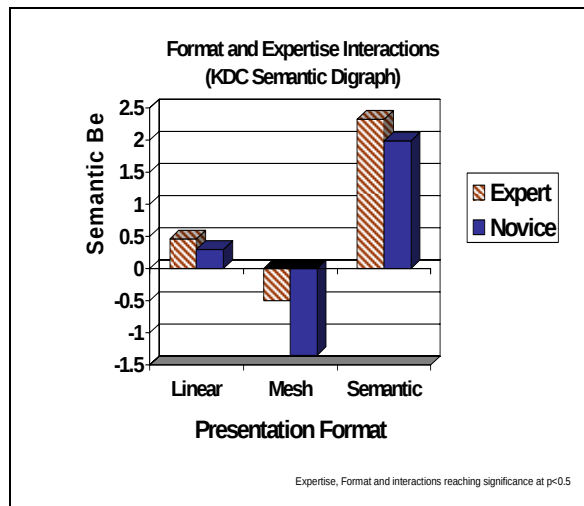


Figure 5: KDC Semantic Digraph analysis suggest that when presented with Semantic-based hypertext format both novice and experts traversed semantically. In addition, experts and, novices to some extent, seem to be moving between web pages (nodes) sequentially in a linear fashion.



These findings demonstrate that the ordering of web-pages visited in the traversal data possess meaningful statistical regularities that can be detected using KDC theory. The qualitative pattern of results for the KDC data analysis were similar to the quantitative results for the ANOVA. In addition, KDC analysis showed a significant trend indicating that participants in the novice group tend to visit web-pages in a linear sequence whereas experts seem comfortable in employing both sequential and semantic-based navigation strategies. In addition, it seems that expertise differences are being minimized when readers are presented with a semantic-based hypertext format. We must note that these trends are not apparent using the classical ANOVA analysis. These preliminary results are encouraging and warrants further studies in exploring more detailed power comparisons KDC and ANOVA data analyses.

It seems that latent expertise knowledge models and web presentation format effects can be discerned by traversal behaviors if presented with discourse inspired web formats using categorical time series analysis.

As reported by Ismaili & Golden (2008) using classical ANOVA analysis that, in general, strong expertise requires more time reading but only for the nodes with most semantic connections and super-ordinate nodes. Novice tend not to spend proportionately more time on these nodes [1]. However, only after further exploration using Knowledge Digraph Contribution (KDC) analysis that we are able to discern underlying navigation patterns and differences between experts and novice [28]. Furthermore, in alignment with past research suggesting navigational paths differences between focused and less focus readers [17], KDC seems to provide confirmation of traversal pattern differences

between expertise and different web presentation formats.

For example, consider comparing 'Mesh' format (middle bars) ANOVA analyses (figure 2) with KDC results for 'Mesh' digraph (see figure 4). ANOVA result suggests that experts' tend to spend significantly more time reading *link* nodes as compared with novice. From KDC it is shown clearly that they did not end up visiting these *link* nodes randomly but rather used sequential (linear) traversal strategies which are not evident in navigation patterns for novice.

Although more work is warranted in this area however the preliminary results using KDC analysis are encouraging in suggesting that novice seems to be more comfortable using linear navigation strategies, whereas experts are shown to be employing both the linear and semantic traversal strategies. In addition, it seems that expertise differences are minimized when readers are presented with a semantic-based Hypertext formats. Recall that semantic-base hypertext format was created using a meticulous content design process both the micro and macro level across all three knowledge domain areas.

This paper seeks to present a framework and methodology by demonstrating the specific characteristics of expert and novice readers as gauged by their traversal patterns in varying web presentation formats. The observed trends are very encouraging however warrants further research. It is hoped that eventually more work on this topic will serve as a crude guide for web instruction designers, human-computer professionals and intelligent web and knowledge engineers to conceptualize or build hyper-learning environments that will not only minimize expertise differences but will also engage experts equally with the learning process.

Moreover, it seems that some of the inconsistent findings from the hypermedia research, in particular, those pertaining to learning systems in web environment may be approached using established findings from discourse literature. Perhaps, moving forward further work in unraveling and discerning the underlying effects may well lie at the nexus of hypermedia and discourse scientific efforts.

6. Acknowledgment

We are very grateful to National Science Foundation (NSF) for their generous support. This project is fully funded by NSF Grant 0624983 from the Methodology, Measurement, and Statistics (MMS) Program in the Division of Social and Economic

Sciences. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References:

- [1] Ismaili, P. B., & Golden, R. M. (2008). Traversal Patterns for Content Designed Web Environment. *WSEAS Transactions on Information Science and Applications*, 11, 5, 1521-1530.
- [2] Kintsch, W., *Comprehension: A paradigm for cognition*, Cambridge University Press, 1998.
- [3] Swaffar, J., Arens, K., & Byrnes, H. (1991). *Reading for meaning: An integrated approach to language learning*. Englewood Cliffs, NJ: Prentice Hall.
- [4] Horiba, Y. (1996). Comprehension processes in L2 reading. *Studies in Second Language Acquisition*, 18, 4, 433-473.
- [5] Halliday, M., & Hasan, R. (1976). Causal coherence and memory for events in narratives. *Cohesion in English*. London: Longman.
- [6] Long, D. L.; Oppy, B. J.; & Seely, M. R. (1997). A "global-coherence" view of event comprehension: Inferential processing as question answering.
- [7] Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of memory and language*, 24, 612-630.
- [8] Rumelhart, D. (1977). Understanding and summarizing brief stories. In D. Laberge & J. Samuels (Eds.), *Computational models of discourse*. Cambridge, MA: MIT Press.
- [9] Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of memory and language*, 24, 612-630.
- [10] Sweller, J., Chandler, P. Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General* 119 (2) 176-192.
- [11] Macedo-Rouet, M., Rouet, J.-F., Epstein, I. & Fayard, P. (2003). Effects of Online Reading on Popular Science Comprehension. *Science Communication* 25 (2), 99-128.
- [12] Foltz, P. W. (1996). Comprehension, coherence, and strategies in hypertext and linear text. In J.-F. Rouet, J. J. Levonen, A. Dillon, & R. J. Spiro (Eds.), *Hypertext and cognition* (pp.109-136).
- [13] Britt, M.A., Rouet, J.-F., & Perfetti, C.A. (1996). Using hypertext to study and reason about historical evidence. In J.-F. Rouet, J. J. Levonen, A. Dillon & R. J. Spiro (Eds.), *Hypertext and cognition* (pp. 43-72). Mahwah, NJ: Lawrence Erlbaum Associates.
- [14] Lawless, K. A., Brown, S. W., Mills, R. & Mayall, H. J. (2003). Knowledge, Interest, Recall and Navigation: A Look at Hypertext Processing. *Journal of Literacy*, 35 (3), 911-934.
- [15] McDonald, S., & Stevenson, R. J., (1998). Navigation in hyperspace: An evaluation of the effects of navigational tools and subject matter expertise on browsing and information retrieval in hypertext. *Interacting with Computers* 10, 129-142.
- [16] Puntambekar, S., Stylianou, A., & Hübscher, R. (2003). Improving navigation and learning in hypertext environments with navigable concept maps. *Human-Computer Interaction*, 18, 395-428.
- [17] Lawless, K. A., & Kulikowich, J. M. (1998). Domain knowledge, interest, and hypertext navigation: a study of individual differences. *Journal of Educational Multimedia and Hypermedia*, 7, 51-70.
- [18] Golden, R. M. (1994). Analysis of categorical time-series text recall data using a connectionist model. *Journal of Biological Systems*, 2, 283-305.
- [19] Sanderson P.M., Benda P.J. (1998). Exploring sequential data: Commentary on Bowers, Jentsch, Salas, and Braun. *Human Factors*, 40, 680-684.
- [20] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1-25.
- [21] Ismaili, P. B. (2008). Can expertise be discerned from traversal behavior in a content designed hypertext (web) environment? In the *Proceedings of the 8th WSEAS International Conference on Distance Learning and Web Engineering*. University of Santander, Cantabria, Spain.
- [22] Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.
- [23] Trabasso, T., Secco, T., & van den Broek, P. (1984). *Causal cohesion and story coherence*. In H. Mandl, N.L. Stein, & T. Trabasso (Eds.) *Learning and comprehension of text*. (pp.83-111). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- [24] Trabasso, T. & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Processes*, 21, 255-287.
- [25] Bower, Black & Turner. (1979). Scripts in Memory for text. *Cognitive Psychology*, 11, 177-220.
- [26] Saremi, H., Q. & Montazer, G. M., (2007). An application of type-2 Fuzzy notions in website structures selection: Utilizing extended TOPSIS

method. *WSEAS Transactions on Computers*, 7, 8-15.

[27] Lee, C-H., Lee, G-G. & Leu, Y., (2007). Analysis on the adaptive Scaffolding Learning Path and learning. *WSEAS Transactions on Information Science & Applications*, 4, 320-330.

[28] Ismaili, P. B. (2009). Any identifiable Knowledge Structures from Traversal Behavior in Hypertext environment? In the *Proceedings of the 8th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*. University of Cambridge, Cambridge, UK.

[29] Golden, R. M. (1998). Knowledge digraph contribution analysis of protocol data. *Discourse Processes*, 25, 179-210.

[30] Bousfield, A. K., & Bousfield, W. A. (1966). Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, 19, 935-942.

[31] Gottman, J., & Roy, A. K. (1990). *Sequential analysis: A guide for behavioral researchers*. New York: Cambridge University Press.

[32] Fredin, E. (1997). Rethinking the news story for the internet: hyperstory prototypes and a model of the user. *Journalism and Mass Communication*, 163, 1-47.

[33] Jacobson, M. J., & Spiro, R. J. (1995). Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: An empirical investigation. *Journal of Educational Computing Research*, 12(4), 301-333.

[34] Dillon, A., & Gabbard, R. (1998). Hypermedia as an educational technology: a review of the quantitative research literature on learner comprehension, control, and style. *Review of Educational Research*, 68(3), 322-349.