Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 159 (2019) 1135–1144

Procedia Computer Science

www.elsevier.com/locate/procedia

23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

On Managing Contextual Knowledge of Digital Document Ecosystems, characterized by Alphanumeric Textual Data

Shastri L Nimmagadda*, Dengya Zhu and Torsten Reiners

School of Management, Curtin University, Perth, WA, Australia

Abstract

The multidisciplinary textual-data are often disorganized and misinterpreted in many documents, which can obscure the information retrieval and its interpretation in company networks and even the World Wide Web. Managing textual information in particular with large-size alphanumeric data sources is challenging and at times can preclude the prompt delivery of good quality document services to diverse customers. Optimizing the words, sentences and alphanumeric characters of a script is the purpose of research, without losing intelligibility, semantics, perception, content flow and the contextual scenarios, represented as dimensions. We interpret the manuscript as a document ecosystem, within which different dimensions are construed. We choose different lexes, sentences, paragraphs and pages that possess frequent alphanumeric characters, interpreted in multiple domains and contexts. The ontologies of alphanumeric textual-data dimensions and their metaphors are presented in several data schemas, connecting various contexts of document ecosystems. The domain ontologies that can deliver text-mining, the semantic and schematic information of textual data, can expedite the textual-data integration process in the multidimensional warehouse modelling procedure. Diverse views and contexts that are generic within the document ecosystems are analysed for contextual knowledge. The ontologically structured document ecosystems that can facilitate more legibility and reproducibility to a variety of document designers are research outcomes. Data analysts, text mining experts and document managers can benefit the current research.

Keywords: Big Data; Data Modelling; Digital Document Ecosystems; Sustainability; Text Mining

1. Introduction

The quality of the document depends on how logically the words and their alphanumeric characters are organized in contextual knowledge domains. The transcripts of documents are enclosed with several conceptual- and contextual-

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: shastri.nimmagadda@curtin.edu.au

1877-0509
10.1016/j.procs.2019.09.282

based semantic, schematic and syntactic expressions [1] in multiple domains. In computer and information science perspectives, the document ecosystems need ontology structuring [2] for making connections between alphanumeric characters and their linked words. We examine entities and or dimensions, their types, and interrelationships between words that may have existed in a specified domain or domains. The ontologies are purposeful artefacts for categorizing or classifying the dimensions, called taxonomies [18]. In the context of digital document ecosystems, we interpret words, sentences, punctuations, grammar, and spell-checked expressions to improvise the connections between conceptualized and contextualized attributes. The quality and integrity of documents is further evaluated, by judging the plagiarism of the textual data. In addition, the retrieval of the desired information from contextual documents has significance on their presentation, visualization and interpretation.

The information retrieval from the World Wide Web is often a tedious and complex process. Data or Information could be from many domains, classes and or contexts. For example, a document manager needs a piece of information from large volumes of textual-data pertained to multiple healthcare, human ecology, human anatomy or food domains. The specific data or information search relevant to the one-to-one domain is not easy from a variety of data sources. For this purpose, we propose a more robust methodology, such as ontologically structured multidimensional textual-data warehousing and mining approach to explore connections in between alphanumeric characters and discover an interpretable knowledge. Data visualization and interpretation are the other two major challenges inherently linked to the implementation of warehoused metadata models. In our approach, alphanumeric characters are ontologically modelled to their atomic levels [15], intelligently integrated and stored in a warehouse environment as unified metadata. Textual data mining and interpreting the words, sentences, text, contexts and phrases relevant to various documents in the industry scenarios are pertinent to artefacts discussed in [19]. Data and or information may be numeric, alphanumeric, integers and floating character instances. Such data routinely input in different types of clerical jobs and applications.

Data and information may be fed from a manual source to an electronic source. Typically the textual data entered in digits are known in advance in addition to their tacit knowledge. Unless stated, there may be a need to change or edit the data, which can't simply be altered or scanned in visual formats. Data that do not need editing can necessarily be made available across multiple systems or databases. The information is shared in a format that's more recognizable to the systems, but not by an image that can provide an ambiguous interpretation. It may be cost effective for data to manage in textual form by processors rather than processing information from complex scanning and optical character recognition systems in abstruse form. We provide innovative ideas of managing the content and styles of the documents, presentable in multiple domains and industry scenarios.

The rest of the research paper is structured in various sections. The definitions of digital document ecosystems, the importance of information retrieval from large size manuscripts and data entry methods are given in Section 1. Various issues and challenges of information retrieval from document ecosystems are given in Section 2. The research objectives and motivation of the research are given in Section 3. Data modelling is described in Section 4. An integrated methodological framework is given in Section 5, accommodating the ontology-based alphanumeric data warehousing and mining articulations. Analysis of results and discussions are given in Section 6. The research is concluded with future scope in Section 7.

1.1. Literature Survey and Identifying Research Gaps

An architecture is presented in [4] that supports the seamless manipulation of Paper Augmented Digital Documents (PADD) with report of “lesson learnt”, implementing the PADD. The authors argue that the PADD is suited to proofreading, editing and annotation of documents like blueprints. New insights of digital agility and transformation of business documents are provided in terms of cost savings and customer satisfaction [7]. Contextual knowledge that can revolutionize the digital transformation in multiple domain applications has not been dealt with by earlier researchers. New tools and technologies are articulated while analysing the digital transformation in multiple contexts including their implementation in different domain applications [11, 12]. The design aspects of models that can retrieve information needed for process analysis is discussed in [20]. So far none of the researchers is discussed the managerial aspects of contextual knowledge while using alphanumeric-based textual data and their associated frameworks. The authors propose new textual-data modelling approach using multiple dimensions.

All the dimensions of the document exhibit tolerance, revealing a balance of semantics between words and their alphanumeric characters. The dimensions are capable of being used, reused, including the document’s validity properties [10]. The document ecosystems may be a single paragraph or a large thesis document. Like any other ecosystems, the document ecosystem varies with geographic and periodic contextual dimensions [3]. The knowledge concealed within the document ecosystems depends on their quality, validity, integrity, usability to ultimately test the effectiveness of the text in a manner the message and or meaning of the document is delivered to the interpreter. The information and the digital textual-data instances of the document ecosystems, how swiftly they can be retrieved depend upon how the dimensions are logically structured in various multidimensional schemas. The keywords and vocabularies [5] used in the document are defined in different tags for text mining [19].

1.2. What is Digital Document Ecosystem?

An active document consists of a user-specified number of pages, words, alphanumeric characters, paragraphs, sentences, with structure, style, punctuation, grammar with contextual spellings, all needed to go through plagiarism process. We view them as entities and or dimensions. The document is characterized as an ecosystem, implying that a manuscript is a complex set of relationships between these dimensions. They vary in size and depend on other related documents, their structures, concepts and contexts. If any part of the text changes, the other parts do change including their semantic, schematic and syntactic content [14, 15]. When a document ecosystem generates an interpretable new knowledge, we say it is sustainable document [6, 8].

The alphanumeric data how they emanate in both research and industry contexts [6] are described in the following sections.

1.3. Data and Information Loaders

Data entry that requires the input of alphanumeric characters into electronic documents in the digital environment is termed as transcription of alphanumeric data in different media. If not regularly, many of the data entry jobs demand alphanumeric work in different contexts and occurrences [3]. It is also significant to understand the words or phrases built based on concepts and contexts or vice versa, for which specific words and characters are considered. The characters could be from various sources of data and information, or even from a variety of online networks or offline research communities.

1.4. Rate of Retrieval

Often, the companies hire hands-on personnel with alphanumeric data entry skills, primarily based upon keystroke speed. Speed is measured in keystrokes per hour rather than words per minute as it is described for regular typing. Most jobs require that alphanumeric data entry operators' type between 7,000 and 12,000 KPH. In addition, the accessibility of a piece of data and information from large volumes of databases in the shortest period of time is another criteria.

2. Significance and Motivation

The current research is aimed at developing different dimensional models, as ontology-based alphanumeric constructs [10] to represent in different contexts. Similar alphanumeric contexts are analysed with a case study as discussed in [14]. In our research, an integrated framework is proposed, accommodating composite dimensions such as alphanumeric characters conscripted from words, sentences and paragraphs in diverse digital document ecosystems. Optimum use of dimensions is the criteria while constructing document ecosystems without sacrificing the content, semantics and knowledge of the document. The significance of the research is to provide constructs, models and innovative methods of optimizing the textual-data dimensions and fact instances of document ecosystems.

3. The Problem Statement and Research Objectives

The manifestation of alphanumeric words is a shared expression of the digital document ecosystems, categorizing various Roman alphabets and numbers. In total, either 36 single cases or 62 case sensitive alphanumeric characters are expressed [1, 14]. But the actual alphanumeric character set in use is numbered from 0 to 9 and alphabets from A to Z (or a to z). While managing the digital document ecosystems, the alphanumeric expressions are at times challenging to organize, and sometimes in large number, they can distract the knowledge management.

A large number of words, alphanumeric characters and sentences of digital documents that have different contexts need large storage devices and technologies to store, process and present documents in various knowledge domains. In relation to computing storage of vocabularies, alphanumeric character significantly stores less than 8 bit ASCII character; each character can only occupy 6 bits in length although there is no standard. A 6-bit character can even have 64 combinations, but 36 are only used in a single case, leaving a room for allowable another 28 characters with punctuations, storing alphanumeric data that contain text and website addresses [18]. If the storage medium is considered in-base 36, the smaller alphanumeric character can even be accommodated in such a storage medium. Type base 36 or 64 storage, having 6 bits per character may be more memory efficient [18, 19] for keeping text-only data, rather than base 256 with 8 bits or a byte. The problems are relevant to (a) an information retrieval from digital document ecosystems and categorization of documents as per contexts (b) developing textual-data models, data warehousing and mining, visualization and interpretation artefacts. These artefacts are feasible for exploiting the contextual knowledge of the text; words and phrases for determining an optimum number of alphanumeric characters to accommodate within sentences and texts logically. Managing the semantics, schematic, syntactic and logical content of contextual scripts is an added challenge while building multidimensional models and knowledge management. Realistically, the alphanumeric representation needs enhancements in textual-data constructs, models and methods and or in the current lexical usage [8], even with subtle morphological variations. Optimizing the lexical and taxonomy dimensions and developing quality digital documents without losing the semantics that can deliver new domain knowledge are key research objectives.

3.1. Ontologies Describing the Document Ecosystems

While describing ecosystems in any document domain, we start with the characterization and description of alphanumeric data sources, from which the alphanumeric-based ontology models need knowledge-based fact instances. Textual data consist of millions of sentences with recurring alphanumeric characters, words, grammar, and punctuation with contextually implied attribute dimensions that can be stored in a single repository, termed as a multidimensional repository [14]. In an analogy, the articulated digital ecosystems consist of several documents, each document system is described, compatible with an information system, with information on several concepts and contexts with diverse and interrelated scenarios of the documents. Each document is characterized by a number of pages, paragraphs, sentences, words and character attributes. It is an initial hierarchy, starting from the generalization to specialization levels, interpreted for each document. It is inherently a simulated digital ecosystem, whose member attributes are hierarchically connected [9]. To demonstrate the phenomenon, we consider several words, characters, sentences analysing various contexts, as described in Table 1.

Table 1. Attribute Dimensions for Contextual Analysis

D	**W**	**C**	**C-S**	**P**	**L**	**Context**	**Type**
1	179	1056	1234	1	15	Petrol	Org
	203	1209	1411	1	17		Edt
2	229	1307	1535	1	18	Health	Org
	283	1642	1924	1	23		Edt
3	284	1569	1853	1	20	Medical	Org
	283	1565	1850	1	20		Edt
4	254	1606	1859	1	21	Food-diabetic	Org
	281	1714	1994	1	41		Edt
5	228	1483	1710	1	19	Ecology	Org
	271	1667	1938	1	42		Edt

D: dimension; W: words; C: characters; C-S: characters with space; P: page; L: sentences; Org: Original document; Edt: Edited document

As it relates to any business, an ecosystem is viewed as an orderly conceptualization and contextualization, where the textual-data relationships established through dissimilar documents become conjointly positive, self-sustaining and fairly closed [9, 13]. It is clearly the case for a broader document ecosystem extended and narrated among several concepts and contexts, as a digital ecosystem, an evolving phenomenon. In this approach, all the existing attributes of document ecosystems are ontologically characterized and interrelated in a warehouse environment. The aim of the ecosystem is to express and generate "local translate to global" documents. Multiple data types may typically characterize textual-data heterogeneity describing:

1. Syntactic heterogeneity – differences in formatting the alphanumeric data.
2. Schematic or structural – structural differences during accumulation and storage of data.
3. Semantic heterogeneity – differences in the interpretation of the *meaning* of alphanumeric textual data from multiple sources and contexts.
4. System heterogeneity – reconciling and accommodating the differences in the operating and hardware systems, keeping in view the differences in formats and dimensions of the document ecosystems [15, 16], [9, 10] and [5].

4. Textual-Data Modelling Methodology

We provide step by step process to create an ontology structure within a given domain:

1. Define the context in a specified domain.
2. Make classes and their hierarchies.
3. Identify the attributes of classes.
4. Connect class attributes with identified inter-relationships.

The alphanumeric characters appeared in different contexts in a given domain or domains are documented. For connecting the alphabetic and numeric letters in various contexts of classes, we need to understand their hierarchies. Accordingly, all the contextual attributes, their respective classes are identified by building inter-relationships among alphanumeric characters. In this context, for building multidimensional ontologies, we estimate the size of the words, the number of alphanumeric characters to typically manage them in the MS office window, as shown in Fig. 1. The Classified Interrelated Contexts of Alphanumeric Ontology (CICAO) is the main data model terminology, as referred to in this context [6].

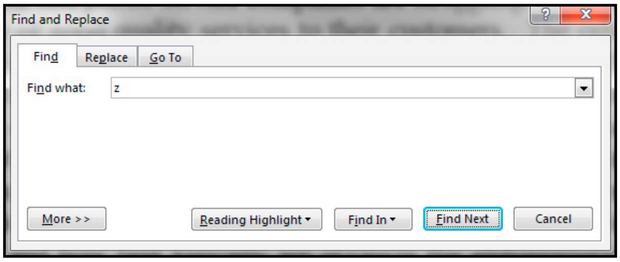

Fig. 1: Editing window, browsed for words and alphanumeric characters

A high degree of granularity is needed with specialized field ontologies [16] in the current contexts. The processes of conceptualization and categorization desirable on classes of concepts, nodular models and diverse conceptual relations are adequately presented in an agreed specific field. Though they are not superior to any documented thesauruses, the hierarchical and associational interactions establish the relationships in pre-established categories. At a theoretical level, ontologies allow specialized concepts in a more open manner, making use of a great range of conceptual relationships [8, 16], as they are not restricted from the outset by a particular practical objective.

An automatic abstract generation, automatic terminology extraction and textual data mining [19] are other examples of new tools. The new search for specific words (information) within texts or sets of texts is still based on concepts and or contexts [3]. For this purpose, ontologically described alphanumeric characters are gathered to integrate with a set of words that make sentences and text within a document. For example, *Year of Oil Spud* and *Petroleum Production in the 2011*, in which there are several alphanumeric characters that make the words and their alphanumeric characters make a meaning, such as "date of spudding the well" and "production of oil & gas in year 2011 is 6000BOPD". Alphabetic characters, such as "d", "a", "t", "e", are captured from the structured metadata (which are either relationally or hierarchically structured) and then integrated them in a warehouse environment to make up "Date" word. The structures of alphanumeric data are favourably made granular, making elements such as

Lexis, lexical combinations and text structure acquire with a renewed textual expression. Additionally, an involvement of linguistic work either in petroleum or healthcare information retrieval is related to the results brought about by taxonomies. Similarly, “P”, “E”, T”, “R”, “O”, “L”, “E”, “U”, “M”, either capital or small alphabetic characters are captured from the ontologically described multidimensional warehoused metadata to make interconnections between alpha-numeric characters, such as “6”, “0”, “0”, “0”, “0”, “B”, “O”, “P”, “D”, that make an expression, as interpreted by oil and gas explorers. As an example, we consider the following paragraph with a set of alphanumeric datasets.

We text mine [19] the paragraphs in terms of a number of sentences, words and alphanumeric characters of each word. We interpret the appearance of the number of alphanumeric characters “a”, “b”, “c”, “d”… in each word, the number of words in each sentence, paragraph and several paragraphs of each document. We try to manipulate the characters in each word in such a manner, we improvise the quality of the text and thus the entire document ecosystem. The dimensions considered in the multidimensional modelling process that connected through schemas are the number of pages, words, paragraphs and sentences at the generalization level, representing the number of alphanumeric characters at specialization level.

As demonstrated in Figs. 2a and 2b, the vowels, from “a” to “u” and consonants from “b” to “z”, are treated as dimensions in the dimensional modelling process. The numbers “0” to “9” are numeric dimension instances incorporated with the alphabetic letters and their schemas. The purpose of star-schemas is to connect the alphanumeric characters in multiple contextual documents. These are generalized schemas, built to work any contextual document.

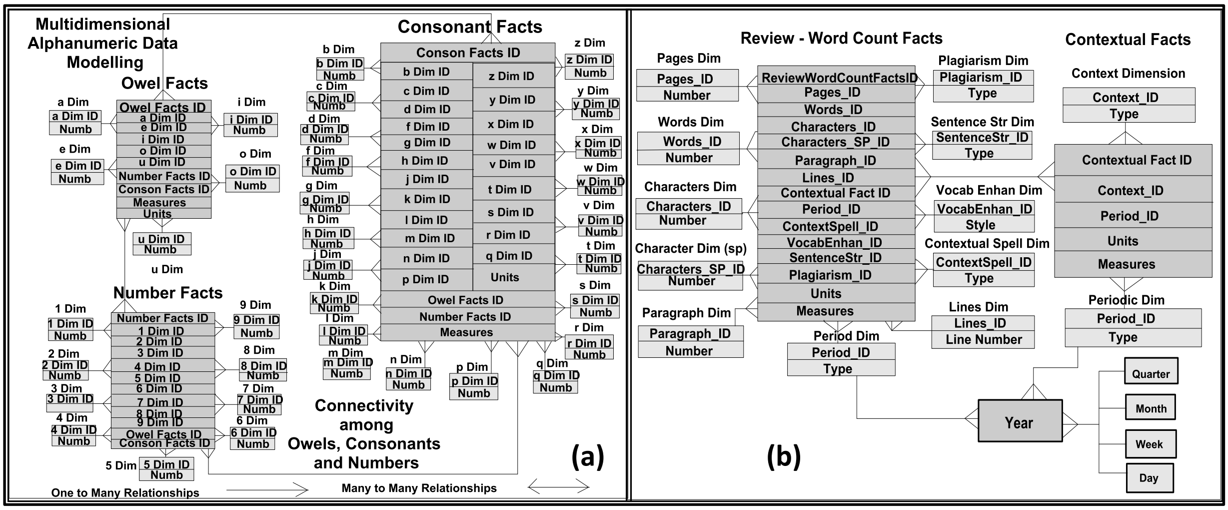

Fig. 2. (a) Modelling the alphanumeric characters and connectivity (b) Schema with word count and contextual facts

In the present study, we consider the healthcare, human ecology, food and medical related domains and petroleum, an unrelated context to test and validate the constructs and models. The petroleum is a non-coexistent domain, not relevant to the other coexistent healthcare, human ecology, food and medical domains. We develop a framework that can accommodate various data structures involving dimensions related to alphanumeric characters and their connectable words as described in Section 5. We test the patterns and trends of metadata views in various plots as presented in Section 6.

5. System Architecture - Integrated Methodological Framework

For addressing connectivity issues and challenges in the problem statement, we design an integrated framework and explore connections between various attribute dimensions of the document ecosystems. As discussed in [9, 10, 13, 15] star, snowflake and fact constellation schemas are commonly adopted in building multidimensional constructs and logical data models. As suggested in [13], the data are hierarchically structured in unique knowledge domains in a warehouse environment. In other words, ontologically structured data are warehoused through the integration of multidimensional data structures. In addition, the data structures designed in the current contexts target the fine-grain structuring of alphanumeric characters so as to follow up the atomic properties while scheming them into words. The process of using/reusing and integrating domain knowledge from fine-grained metadata is another significant part of the document ecosystem methodological framework.

We correlate the ontologies in the current application scenarios with taxonomic hierarchies of classes, class definitions, and class conceptualizations that emerge while building textual-data relationships in multiple dimensions [18, 19]. To identify the conceptualizations, business rules and axiom constraints need to be committed during contextual interpretations of the conceptualizations. In the context of a unified workflow (Fig. 3), the concept of an ecosystem is beneficial with number of associative entities or dimensions' presence in the integration process through symbiotic positive sum interactions between words [9]. It is a complex relationship with number of multiple dimensions and its environment is functional as an ecological unit. It is an analogy, volume of attributes and their instances are gathered and characterized by multiple data sources, all stored in a single repository. The similar comparison is made in a broader sense of a large-sized document file, comprising of multiple document ecosystems with several hundreds of attribute dimensions and instances. They are connected to other large-size contextual documents in an inclusive manner, where alphanumeric characters' data do have semantic and syntactic boundaries. Various tools such as thesaurus, word validator and checker, domain/context including grammar checker are connected to multidimensional schemas and ontology based warehousing and mining architectures to make the documents semantic, schematic and syntactic.

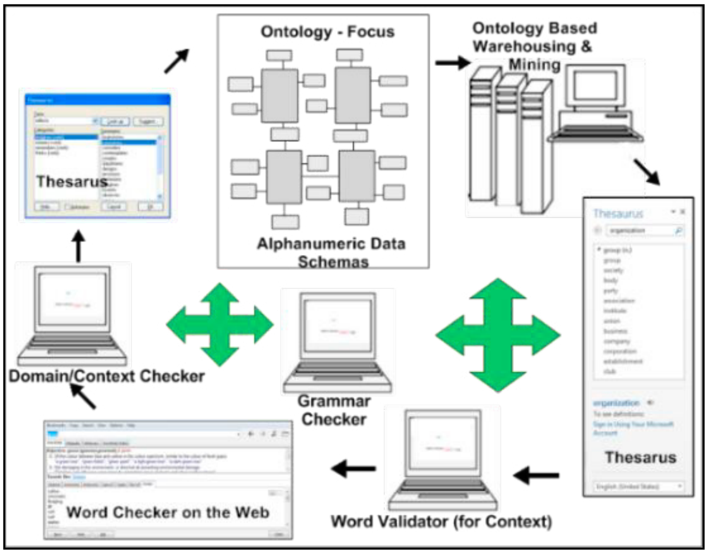

Fig. 3. A framework for ontology-based alphanumeric data warehousing and mining

Intelligent and expert data systems are already in place in industry scenarios with motivated smart computer simulations, demonstrating their applicability and feasibility and evaluating the contextual knowledge of digital document ecosystems [17]. We test the patterns and trends of alphanumeric characters and their linked words with their contextual instances. They are interpreted and presented in various scalar line plot views in the following sections.

6. Results and Discussions

Having ascertained the research gaps with description of significance and motivation of the research, we explain the research contribution in this section. The evaluation of framework and beneficiaries of the research are discussed in the following sections.

6.1. Evaluation, Value of Research and Beneficiaries

The research is validated through various artefacts of the document ecosystems adaptable in multiple domains. While drafting the manuscript, several criteria and guidelines are followed for improvising the quality of the content to make a readable and meaningful script. Improvising the “contextual spelling”, “grammar (such as appropriate use of propositions, articles and verbs)”, “punctuation”, “sentence structure”, “style of expression (such as voice)”, “vocabulary enhancement” including the plagiarism and an overall score of the document are the current focus. The framework uses a number of contextual words, composing several alphanumeric characters including spelling, grammar, vocabulary enhancements and plagiarism checks done. As described in Table 2, several such contexts are analyzed, and the error status and score of the documents are tabulated. As shown in Figs. 4a and 4b, bubble plots are drawn showing the strength and appearance of vowels and consonants in the contextualized words. The vowels and consonants display positive trends, distinctly larger bubbles with the vowels and smaller bubbles with the consonants.

Table 2: Interpretation of Attribute Dimensions and Their Qualities

D	CP	G	P	SS	S	VE	P	S (%)
1	1	0	0	0	0	1	0	97
2	0	0	0	0	0	0	0	100
3	1	0	0	1	0	0	0	98
4	2	0	0	2	1	0	0	95
5	3	1	0	1	0	0	0	97

D: document; CP: contextual spelling; G: grammar; P: punctuation; SS: sentence structure; S: style; VE: vocabulary enhancement; P: plagiarism; S: score (%).

Another criterion is, designing the document with an optimum number of sentences, words and alphanumeric characters, without compromising the semantics and contextual interpretation of multiple words in new knowledge domains, maintaining the overall quality of the document [18]. The bubble plots can exhibit patterns of dependent or independent variables in different scalar descriptions. In a 2D bubble plot, the diameter of each bubble varies in size, presenting a way to characterize new dimensions of data in multiple contexts. The impacts of vowels and consonants in different contexts are shown in Fig. 4a, suggesting a stronger existence of vowels in many contexts compared with the presence of consonants in their smaller bubble sizes with lesser influence of contexts. Another interesting observation is that encircled lobes and clusters suggest “healthcare”, “food”, “human ecology” and “medical” contexts and their alphanumeric characters are closely related, and their clustered links are interpreted in the coexistent domains.

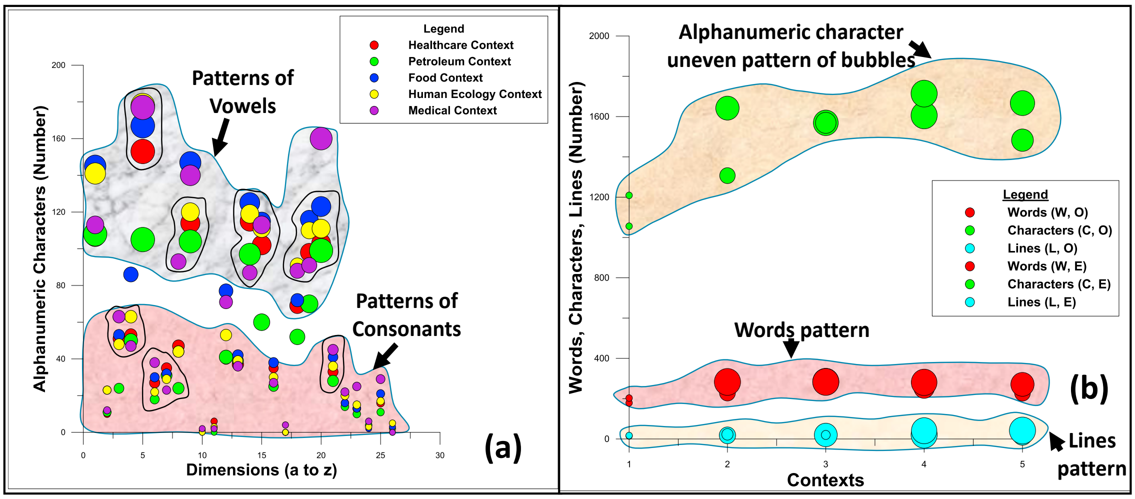

Fig. 4. (a) Bubble plot view of the effect of vowels and consonants in various contexts and clusters (b) Contextual representation of words, characters and sentences

In another display of the contextual interpretation of alphanumeric characters, words and sentences of the document files, more occurrence of alphanumeric characters is observed with uneven patterns in different contexts. Whereas words and sentences that represent small bubble sizes are effective in the interpretation of the contextual knowledge after editing them, as shown in Fig. 4b. For example, words (O, original): words (E, edited); sentences (O, original) and sentences (E, edited) that composed of alphanumeric character dimensions are compared. The edited and original patterns of alphanumeric characters show uneven trends, whereas patterns attributed by words and scalar line plots exhibit even and coherent shapes of bubbles as shown in Fig. 4b. A similar observation is made with the line plots (the data points are connected by lines) between vowel and consonant dimensions, in which evenly distributed alphanumeric characters are interpreted that make sense of contexts.

The alphanumeric characters such as vowels, consonants and numbers are often used in combination in words in various contexts. At places, lower or upper cases are used based on the contexts and format. These contexts and their use/reuse determine whether or not case sensitivity is applied based on the other words associated with the sentences that may belong to other domains and or document ecosystems.

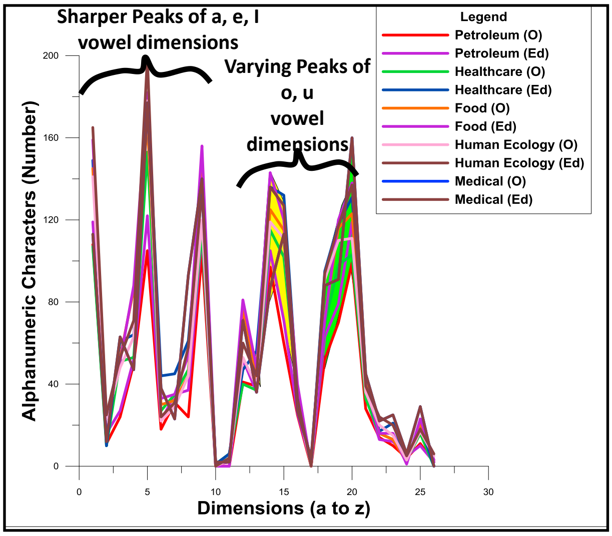

Fig. 5: Scalar line plot view of alphanumeric characters with vowels and consonants dimensions for different contexts

We compare the patterns of vowels of “a”, “e” and “i” with trends of an uneven distribution of “o” and “u” in all words representing multiple contexts as shown in a line scalar plot in Fig. 5. The early patterns of alphanumeric dimensions and their instances of vowels display shaper peaks for the same contexts. Whereas “o” and “u” vowels characterize with wider patterns of textual-data in different contexts. Similarly, coexistent, relevant and unrelated contexts present vertical sharper resolutions of events, suggesting an optimum number of vowel distributions in every contextualized text that incorporated in the manuscript.

The beneficiaries of the research are document designers, who use word processors, academic institutions, and digital transformation researchers working in businesses and marketing and companies investing in digital technologies.

7. Conclusions, Limitations and Recommendations

The framework aimed at integrating and analysing the domain ontologies of the alphanumeric contextual data in a warehouse environment is efficient for granular data mining and new knowledge interpretation. The methodology is effective in unifying the alphanumeric attribute dimensions and representing them in diverse contexts. The procedure can revise the sentences, words and characters used in the construction of various paragraphs of the document ecosystems. The bubble and line-scatter plots drawn between dimensions and their instances are useful in evaluating the alphanumeric characters in contextualized words or phrases. The procedure facilitates extracting hidden knowledge of the scripts distinctly specified in various paragraphs from diverse contexts.

8. Future Outlook

For enduring the competitive markets in the distributed business environment, we emphasize the digital transformation as future scope and opportunity. Businesses done in supply chain environment need large amount of data and information to be represented in multimedia. In academic environment, the current approach is feasible especially interpretation of textual data and information in multiple domains, as manipulations looked-for in large size documents are justifiable.

Acknowledgements

We acknowledge the financial support provided by the Commonwealth Government of Australia for undertaking the current research. We are thankful to the Head of School of Management for permitting us to present our research in the 23rd Annual KES International Conference of Knowledge-Based and Intelligent Information & Engineering Systems held in Budapest, Hungary.

References

[1] Alemi, A.A. and Ginsparg, P. (2015) Text Segmentation based on Semantic Word Embeddings, KDD '15 Sydney, Australia.

[2] Brank, J., Grobelnik, M. and Mladenic, D. (2005) A survey of ontology evaluation techniques, proceedings of data mining and data warehouses (SiKDD); http://www.slideshare.net/ontoini/icist-2013-presentation.

[3] Daquino, M. (2014) Historical Context Ontology (HiCO), http://hico.sourceforge.net/index.html, VERSION 1, 2014.

[4] Guimbretière, F. Paper Augmented Digital Documents, https://www.cs.cornell.edu/~francois/Papers/UIST03.pdf. UIST'05, October 23–27, 2005, Seattle, Washington, USA.

[5] Harpring, P. (2010) Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works, Getty Publications, 13 Apr. 2010 - Art - 258 pages.

[6] Hepp, M. and Radinger, A. (2003) eClass OWL – The Web Ontology for Products and Services, OWL Representation of the eCl@ss Classification Standard, http://www.heppnetz.de/projects/eclassowl/.

[7] Heuvel, B. V. D. (2018) Transform business processes with electronic and digital signatures, Electronic and Digital Signatures in Adobe Sign, White Paper, 2018.

[8] Lacasta, J. Nogueras-Iso, J. Francisco, J. and Soria, Z. (2010) Terminological Ontologies: Design, Management and Practical Applications, Springer Science & Business Media, 3 Aug. 2010 - Computers - 198 pages.

[9] Nimmagadda, S. L. and Dreher, H. (2012) "On new emerging concepts of Petroleum Digital Ecosystem (PDE)", Journal *WIREs Data Mining Knowledge Discovery,* 2012, 2: 457–475 doi: 10.1002/widm.1070.

[10] Nimmagadda, S.L. (2015) Data Warehousing for Mining of Heterogeneous and Multidimensional Data Sources, Verlag Publisher, Scholar Press, OmniScriptum GMBH & CO. KG, p. 1-657, Germany.

[11] Nimmagadda, S. L., Dreher, H.V. and Rudra, A. (2016) "On a Holistic Modelling Approach for Managing Carbon Emission Ecosystems", *Journal of Environmental Modelling & Assessment*, Springer Nature, 2016.

[12] Nimmagadda, S.L, Torsten Reiners, T., and Lincoln C. Wood, L. C. (2018) "On big data-guided upstream business research and its knowledge management", *Journal of Business Research*, Elsevier Publishers, 2018.

[13] Nimmagadda, S.L. and Rudra, A. (2017) Big Data Information Systems for Managing Embedded Digital Ecosystems (EDE), a book chapter in a book entitled "*Big Data and Learning Analytics in Higher Education: Current Theory and Practice",* Springer International, DOI: 10.1007/978-3-319-06520-5, ISBN: 978-3-319-06519-9, The Netherlands.

[14] Protondo, A. (2011) Representation for alphanumeric data type based on space and speed case study: student ID of X University, *International Journal of Database Management Systems (IJDMS)*, Vol.3, No.3, August 2011.

[15] Rudra, A. and Nimmagadda, S.L. (2005) Roles of multidimensionality and granularity in data mining of warehoused Australian resources data, *Proceedings of the 38th Hawaii International Conference on Information System Sciences*, Hawaii, USA.

[16] Shah, M. Representation of ontology using classified interrelated object model, CSE-IT Dept., Nirma University, http://www.slideshare.net/mihika_shah/representation-of-ontology-by-classified-interrelated-object-model.

[17] Sourav, S. B., Josef Küng, J. Wagner, R. (2009) Database and Expert Systems Applications: 20th International Conference, DEXA 2009, Linz, Austria, August 31 - September 4, 2009, Proceedings, Springer, 25 Aug. 2009 - Computers - 865 pages.

[18] Srivastava, A., and Sahami. M. (2009) *Text Mining: Classification, Clustering, and Applications*. Boca Raton, FL: CRC Press. ISBN 978-1-4200-5940-3.

[19] Witten, I.H. (2000) Adaptive Text Mining: Inferring Structure from Sequences, J. of Discrete Algorithms, Vol. 0 No. 0, 0000, Hermes Science Publications, http://www.cs.waikato.ac.nz/~ihw/papers/01IHW-Adaptivetextmining.pdf.

[20] Yen, S.J. (2001) Capturing multimodal design activities in support of information retrieval and process analysis, PhD thesis, Stanford University.