PalArch's Journal of Archaeology of Egypt / Egyptology

KNOWLEDGE GRAPH BONDING WITH THEMES AND CHARACTERS: AN EDUCATIONAL DATA MINING STUDY

Zafar Ullah[1], Muhammad Farooq Aalam[2]

[1]Virtual University, Islamabad, Pakistan,

[2]Assistant Professor, NUML, Rawalpindi, Pakistan;

Email: [1]zafarullah@vu.edu.pk

Zafar Ullah, Farooq Aalam. Knowledge Graph Bonding with Themes and Characters: An Educational Data Mining Study -- Palarch's Journal of Archaeology of Egypt/Egyptology 19(1), 1075-1093. ISSN 1567-214x

Keywords: Knowledge Graphs, Links Tool, Ontology, Educational Data Mining, Linked Open Data

ABSTRACT

A close reading of big data is a time-consuming activity, then finding the interrelationship of themes and characters is a more challenging task for social science learners. To address these challenges, Links tool draws knowledge graphs that show linked open data as a precise visual. These knowledge graphs look like multipronged links in the Educational Data Mining process, and they are interactive in nature to fulfil the needs of digital natives. Links tool constructs knowledge graphs without any programming skill; hence, this data visualization has become very beneficial for learners and teachers during the pedagogical phase and knowledge discovery. This study is an experiment of Links tool on five biographical essays. The current study employs mixed methods for data generation; Connectivism Theory and Hermeneutica Theory have been applied during data analysis. Furthermore, zooming and filtering techniques have also been applied to comprehend the true hermeneutic meanings of Knowledge Graphs. Major findings of this study reveal that knowledge patterns have been extracted in the form of a correct drawing of knowledge graphs, thus, they serve digital hermeneutic purposes. Moreover, bonding of characters and major events show 88.88% accuracy when compared with the source text; hence they reveal epistemological and ontological characteristics.

INTRODUCTION

A knowledge graph (KG) is a brain like a multipronged visual description of linked data and the manifestation of entity-relationship for semantic queries. So, a KG must have epistemological and ontological attributes in its patterns. They reveal unambiguous and identifiable nodes to derive knowledge patterns. All types of KGs have their origin in neuroscience. Precisely, KGs can be defined

as "A graph, composed of a set of assertions (edges labelled in relation) that are expressed between entities (vertices) where the meaning of the graph is encoded in its structure, the relations and entities are unambiguously identified, a limited set of relations are used to label the edges, and the graph encodes the provenance, especially justification and attribution, of the assertion" (McCusker, Erickson, Chastain, Rashid, Weerawansa, & McGuinness, 2018).

KGs resemble Artificial Neural Networks (ANNs) which emphasise the "study of connectivism" (Gurney, 2014, p.17) of characters and key motifs of any text. Similarly, the wiring of neurons has been visualized in knowledge graphs to show the interconnectivity of key themes and characters from the selected biographical essays. In the same way, Google graphs have got international recognition in research, academia and industry and they express linear associations with other applicable variables (Barahate, 2012, p. 13). The psychological and cognitive underpinning lies in the schema which is the cognitive process to interlink categories of information in the brain. As the philosophical foundation is concerned, the bonding between KG and ontology (Ehrlinger & Wöß, 2016) has popularized it in academia because academia promotes the quest for new patterns of knowledge as knowledge discovery does from unstructured data.

As dispersed data in digitized form are increasing day by day and to explore knowledge patterns from terabytes data has become a great challenge. To cope with this challenge of big digitized data, close reading has failed because of the shortage of time and energy. Furthermore, it is a grand challenge to find a unifying force of any piece of writing and to show its implicit graphic pattern from the source text. The second problem is most of social science learners are unable to perform complex machine learning methods to extract knowledge patterns. The third problem arises whether the generated knowledge graph is laden with knowledge or it is just a beautiful visual. So, the current study has been conducted to solve these three problems simultaneously. This study opts for Voyant text mining suite for the drawing of Knowledge Graphs from the source text of biographical essays taught at the intermediate level in Punjab, Pakistan. This tool mines big data texts and small texts in a very short time. Then accuracy and validity of generated knowledge graphs have been calculated by comparing knowledge graphs with the source text, and it has been presented in table 1 in the results and discussion section. The accuracy and user-friendly qualities make this tool reliable and effective for educational data mining.

Learning design is a strategy and act of formulating new practices, digital tools and activities to accomplish educational objectives in the given situations (Mor, Craft, 2012, p. 86). Links tool and knowledge graphs have become learning designs for text mining and understanding the interrelationship of various themes and characters in the selected biographical essays.

We have been motivated to knowledge graphs because Google knowledge graphs are in vogue in academia. Moreover, knowledge graphs look like neurons of brain and human beings extract hermeneutic patterns from them. Thus, new dimensions of texts are explored by linking different nodes with each

other. As a result, this study builds a strong nexus between AI and text for building better and reliable knowledge patterns.

The significance of KGs is manifold that they interconnect relevant key themes from sources of huge texts. Furthermore, it is a new way of knowledge discovery without programming skills. The distant reading technique of knowledge graph develops the bonding between key themes and characters to discuss and interpret the text. In hermeneutics, different layers of knowledge are interpreted by interconnecting different key themes. This time-consuming, highly intellectual work and error-prone work has been done with the use of KGs created by Links tool in Voyant suite. Moreover, human brains can ignore several nodes of entity-relationship; however, the automated Links tool cannot leave any possible thematic link from the text. Another view is that close reading of a text has its own literary pleasure, but "digital natives" need automated digital tools to discover fast and authentic knowledge patterns. Furthermore, NASA and Google also use KGs to elaborate their data and findings. This trend is also very common to study medical repositories for instance PubMed has been studied with AliBaba tool (Plake, Schiemann, Pankalla, Hakenberg, & Leser, 2006). Furthermore, industry and academia have gotten extensive benefits from knowledge graphs. Libraries get ample advantage to explore interrelationships between topics and themes from big data texts. Moreover, recently web-based KGs have also been introduced to extract knowledge patterns from web data, so it mines data at run time. This paper reports on the application and analysis of Links tool from Voyant text mining suite which creates KGs from five selected biographical essays, therefore, it focuses on the creation of KGs and then tests the ratio of accuracy and inaccuracy of KGs by comparing them with source text information.

The first section introduces knowledge graphs and raises intellectual queries establishing its widespread significance. The second section reviews relevant literature on knowledge graphs and tools which produce knowledge graphs. The third section elaborates research methods and their rationales. The fourth section presents results and discussion with data visualizations of knowledge graphs. The fifth section, highlights major findings, implications and recommendations for future research.

Related Literature In Knowledge Graphs

Chronologically, previous studies on knowledge graphs and other relevant works have been reviewed in this section to establish a premise for the current study. KGs are also known with some other relevant terms for example connectionist systems, mind maps, concept maps, neural networks, entity-relationship, linked open data and parallel distributed processing systems. Another term "computational neuroscience" (Churchland & Sejnowski, 1992) utilised mathematics, and computational models to explore knowledge patterns from the domain of neuroscience. Consequently, its KGs were constructed by multiplying inputs with statistical weight.

Later on, the KG construction shifted from text to web-based data. So, a semantic search engine named NAGA was built to create the interconnectivity

of numerous web KGs. They showed more key terms and entity relationships with the increased input of data. To assess its accuracy, a different scoring model was proposed with a generative language model (Kasneci, Suchanek, Ifrim, Ramanath, & Weikum, 2008). As compared to the current Links tool and Hermeneutica Theory study which suggested that KGs "can be extended to expose new things" (Rockwell & Sinclair, 2016, p. 166), but the scoring ability of KGs was lacking in Links tool. To overcome this deficiency, the current study proposed a scoring method of KGs by comparing KGs with the source text.

Another research was conducted to explore and rank the relationship between an entity and a query (Blanco & Zaragoza, 2010), while Voskarides, Meij, Tsagkias, De Rijke and Weerkamp (2015) searched interconnectivity of two terms or themes. One more relevant study was carried out by Fang et al. (2011), consequently, they found a pairing of different entities. In the following year, Baalen (2012) explored mutual connectivism among first persons, second persons and co-occurring terms from songs data. He used Links tool to show KGs but the current study is different and up to date because it uses the latest version of Links tool with modified data visualization interface.

In this digital age, repositories, databases, machine learning libraries, extensive sources of knowledge are very common but extracting meaningful and linked data from big data is a big challenge for common readers, social scientists and educationists. To address this challenge, Sar-graph, a semantic knowledge graph, was built for the word "marriage" by deriving data from Verb-Net, BabelNet, UWN (University World News) WordNet. Consequently, Sar-graph successfully designed an accurate KG of various entities after refining the relevant arguments (Uszkoreit & Xu, 2013).

A systematic literature review regarding educational data mining and learning analytics had been done. Experimental studies had been reviewed from 2008 to 2013. Non-statistical methods with perspectives of pedagogy and technology were applied to 209 works to study their research questions, methods and findings (Papamitsiou, & Economides, 2014).

KGs and corpora were compared for scoring their accuracy. High level of accuracy proved high quality of KG construction (Voskarides, Meij, Tsagkias, De Rijke, & Weerkamp, 2015). The current study also validated or refuted the accuracy of KGs with the help of taking quotes from the text. Then their percentages were calculated to validate the efficacy of KGs. When two corpora were juxtaposed for comparison and verification of discovered knowledge, this method validated Hermeneutica Theory which emphasized to "supplement by other materials" (Rockwell & Sinclair, 2016, p. 166).

When a person tried to discover knowledge patterns from voluminous texts, two issues emerged: the first issue was about the correct relationship of motifs and their elucidation, whereas the second issue was about the relevant pairing of entities and their relatedness. To solve these problems, RECAP tool was introduced to construct correct KGs (Pirrò, 2015) and relatedness of their entities.

In another study, certain KGs were constructed from Star Trek science-fiction movies to analyse characters with a KG. To create KGs, Statistical Relational Learning (SRL) methods were employed but their major deficiency was the absence of common sense (Nickel, Murphy, Tresp, & Gabrilovich, 2016) which caused difficulties in the comprehension of entity-relationship. Sometimes machine learning tools exhibited the absurdities of the absence of common sense.

Machine learning tools never showed a hundred per cent accuracy; hence, its 90% accuracy was considered appropriate. Though several other tools for instance Freebase, Microsoft Satori, Wikidata and Cyc were employed to construct text-based KGs, yet no tool could produce absolutely correct KGs. Later on, web-based KGs were created with Yago and DBpedia. In addition to them, PROSPERA, NELL, KnowledgeVault tools also produced correct knowledge graphs from semi-structured data, and this accuracy was possible by the utilization of heuristic methods (Bordes, & Gabrilovich, 2014 as cited in Paulheim, 2017). Heuristic methods were problem-solving techniques through the use of knowledge graphs.

Another use of KG was introduced to present literary stylistic qualities (Tweedie, Singh. & Holmes, 1996). Following the same research direction, one more stylistic research was conducted on three literary figures for instance Middleton's 90 works, Jonson's 164 samples and Shakespeare's 168 selected samples; so, they were studied with Cascade-Correlation network architecture. It found that Shakespeare was the bona fide author of Thomas Kyd's *'The Spanish Tragedy',* and Madison was the writer who originally penned twelve disputed Federalist papers (Waugh, Adams, & Tweedie, 2000). Consequently, KGs identified the original author of the disputed texts through data mining of stylistic qualities.

Similarly, another author identification study was conducted with the KGs. During cognitive analysis, readers could easily perceive graphic representation and judge textual trends. It decreased the inferential process and enhanced pattern association of text for readers (Shah, Mayer, & Hegarty, 1999). Another study took 89 Pre-service teachers' opinions about the utility of concept maps, and each teacher designed five concept maps; and shared them with other teachers. Their midterm exams, questionnaires presented positive results in the favour of concept mapping for learning and teaching purposes (Koc, 2012).

The Rainbow diagram had been used in the experimental study to assess the interconnectivity of themes in the essay. Then they compared their rainbow graph with a high standard essay to improve it. All participants of higher education opined that OpenEssayist web application was a useful feedback data visualization tool; therefore, users' second draft was better than their first one (Whitelock, Field, Pulman, Richardson, & Van Labeke, 2014). This academic and pedagogic use of KGs facilitated the improvement of academic content. Similarly, the use of Knowledge graphs for other domains had become popular and beneficial. SemaTyP, a time-saving knowledge graph generation method was used in the drug discovery process. After creating knowledge graphs, a logistic regression model was trained, and then a learned model was applied to

find drugs for certain diseases (Sang, Yang, Wang, Liu, Lin, & Wang, 2018). Another network graph analysis study was conducted as a survey study in which 3550 users participated. It emphasized a strong relationship between learning analytics and learning design (Ifenthaler, Gibson, & Dobozy, 2018).

A modern study was designed to introduce an automated KnowEdu system for pedagogical purposes of mathematics and other subjects (Chen, Lu, Zheng, Chen, & Yang, 2018). Furthermore, some other KG tools were designed for instance BabelNet, ConceptNet, DBPedia (built extensive KGs from Wikipedia sources), DeepDive, PROSPERA, ReVerb found binary relations of text and GeoLink tools constructed KGs from the domain of Geography. BioPortal was a web-based tool for finding biomedical ontologies (McCusker, Erickson, Chastain, Rashid, Weerawarana, & McGuinness, 2018).

Some other graphic tools Venn mind maps, Venn diagrams and concept maps were used for knowledge representation (Buitrago, & Chiappe, 2019), but they were not so interactive and well connected like knowledge graphs generated with Links tool. In another contemporary study, Ozymandias was a biodiversity knowledge graph that was based on The Atlas of Living Australia. It visualized the classification of various animals (Page, 2019). Classification and taxonomy constructed knowledge with the aid of knowledge graphs.

In Chinese academic settings. KGs had been employed in academia as a research tool since 2007. The research applied KGs to the Chinese Social Science Citation Index of 2019-2020. The study presented the previous 15 years of works that utilized KGs in education. It revealed trends and topics through Keyword Cooccurrence Network, clustering and time zone map techniques (Mao, 2021).

Automatic answering to questions systems like Siri are increasing and named entity disambiguation are big challenges. This study recommends to expand knowledge graphs with WordNet for better results. Coherence between nodes of KGs should be enhanced. Their similarity should be calculated. Consequently, this study outperformed and showed 27% more accurate results in F-measure (Bouarroudj, Boufaida, & Bellatreche, 2022).

THE RESEARCH QUESTIONS

From the perspective of the research challenges, the current study addresses the following two research queries:

RQ 1. How do Links tool-generated knowledge graphs visualize epistemological and ontological interrelationships of various themes and characters in selected biographical essays?
RQ 2. To what extent do the generated knowledge graphs show accuracy or inaccuracy by verifying or refuting them with the textual evidence of selected biographical essays?

Here epistemology reveals knowledge, and ontology discusses the reality of knowledge and text for hermeneutics. To answer these questions, the current study employs the mixed methods approach but qualitative data were found

more than quantitative data. Its outcome is the drawing of knowledge graphs from textual data of certain literary essays and measurement of their accuracy.

METHODS

The current study followed Saunders, Lewis, & Thornhill's (2012) research onion (figure 1) that covered the five steps: research philosophy, research approach, research strategies, time horizons and data collection methods.

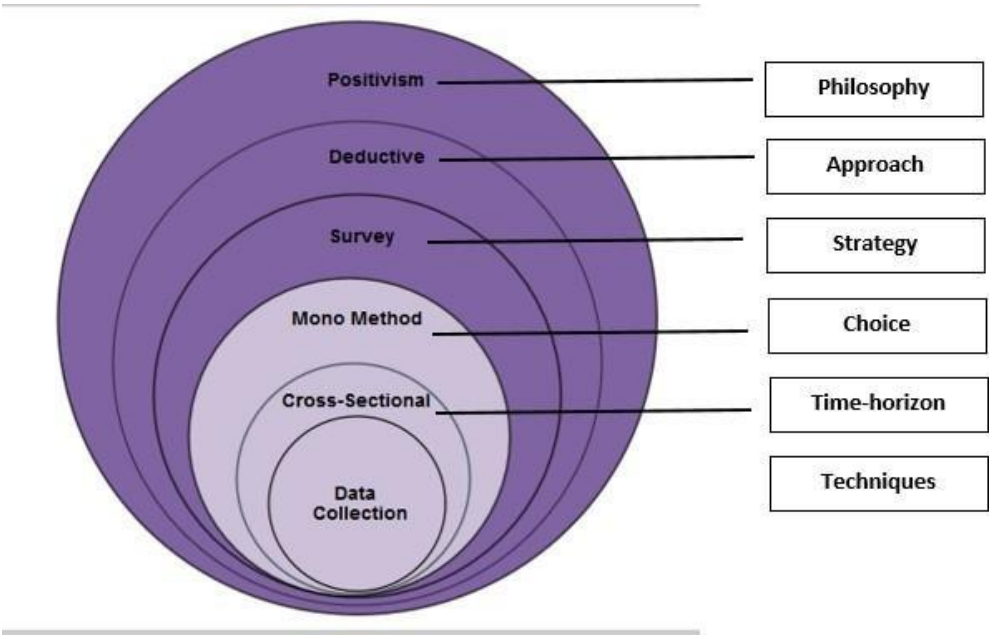

Figure 1: Research Onion (Paulsen, Perera, & Kaddoura, 2020).

Research Philosophy:

The theoretical underpinning of the current study was Connectivism, a conceptual framework, which was propounded first time by Siemens (2005) in his seminal work *'Connectivism: A learning theory for the digital age'*. The learning process occurred in a network managed by socialization and technology. Later Downes (2006) elaborated this theory. Some significant principles of Connectivism were also propounded by Siemens (2005) in the following points:

- "Learning is a process of connecting specialised nodes or information sources.
- Learning may reside in non-human appliances.
- Nurturing and maintaining connections between fields, ideas, and concepts is a core skill.
- Accurate, up-to-date knowledge is the aim of all connectivist learning activities" (Siemens, 2005; Goldie, 2016).

Another philosophical underpinning, Hermeneutica Theory guides that "Manipulation is in service of exploration and understanding." "Knowledge bearing tools provoke reflection" (Rockwell, & Sinclair, 2016, p. 166), especially in the interpretation of knowledge graphs. All bonding of KGs had

been discussed through the lens of connectivism theory (Siemens, 2005) and Hermeneutica Theory. If a KG was in harmony with the source text, it meant that KG was accurate. If a KG differed from the textual information, it was considered inaccurate. The accuracy and inaccuracy of KGs had been counted and presented in table 1 in the results and discussion section. Moreover, its percentage had also been calculated to validate the accuracy of Links tool.

Research Approach:

Apart from it, the current study opted exploratory approach since the construction of KGs explored the ontological and epistemological relationships among textual entities.

Research Strategies:

The research strategy of the current study was to generate knowledge graphs by the use of Links tool from Voyant open access suite. There is a rationale to extract connectivism through tools because "Learning may reside in non-human appliances" (Siemens, 2006). Therefore, the digitized text of each essay was uploaded on Voyant suite to create KGs. Voyant suite was designed by Stefan Sinclair and Geoffrey Rockwell in Canada (Sinclair & Rockwell, 2016). After the creation of KGs, a PNG image of each KG was exported to a Word file for data analysis. To present correct analysis, human beings had to apply critical thinking, filtering, zooming and manipulation techniques on KGs. Likewise, Hermeneutica Theory informs that "Knowledge bearing tools provoke reflection" (Rockwell & Sinclair, 2016, p. 166) because they manifest epistemological and ontological characteristics. Then pairing of KGs was written and its accuracy was judged by the textual quotes.

Time Horizons:

Research data about biographical essays were compiled and digitized during 2021.

Data Collection Methods:

The literary genre was delimited to biographical essays, and they had been taken from intermediate English textbooks taught in Punjab boards of Pakistan. The current study selected five biographical essays about Winston Churchill, Christopher, a hitchhiker, Alexander Fleming, Louis Pasteur and Mustafa Kamal. The rationale behind their selection was that they covered biographies of two scientific personalities, two political personalities and one adventurous hitchhiker. Moreover, they were significant in Punjab, Pakistan because every year more than 100,000 students study these biographical essays for their intermediate exams. The paperbound essays were typed and digitized for the current research.

The current study is valid and reliable because every time the same KG was built from the same textual data. Moreover, Links tool in Voyant suite is also reliable because it is being used as an academic tool in 22 renowned universities

of the world (Sinclair & Rockwell, 2017). Furthermore, it was used for the distant reading process, and it had become an academic and pedagogical need of individuals when they faced big textual data. Human cognition was also involved in the data analysis process because techniques of filtering and zooming of nodes were also applied to derive meaningful patterns. Human brains also performed the same process to zoom and filter themes to present meaningful connections of ideas. Besides, the main research limitation was the loss of interactivity of KGs due to its export on MS word file.

RESULTS AND DISCUSSION

Regarding the first research question, the selected five biographical essays have been analysed along with their knowledge graphs (KGs). Then epistemological and ontological values have been testified with textual evidence. In the concluding part, their accuracy or inaccuracy has been measured in percentage. Final discussion is based on table 1 and figure 7.

First Year at Harrow by Sir Winston S. Churchill

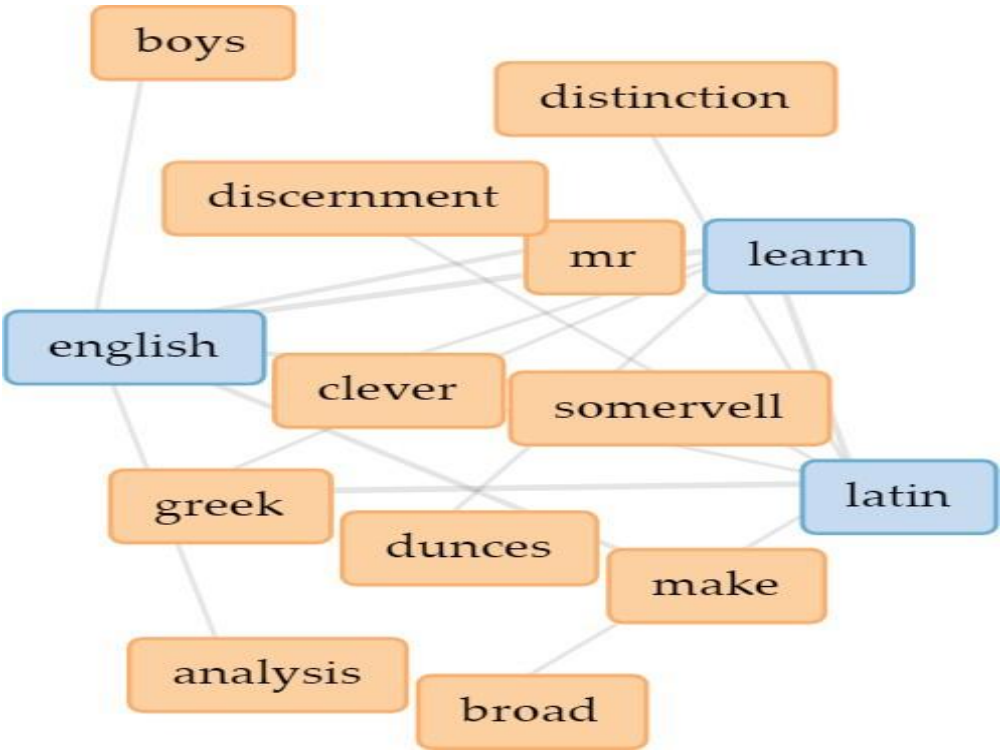

Figure 2 KG of First Year at Harrow

The blue-coloured nodes are “learn” and “Latin” because they are central and have been repeated frequently in this KG. That is why other nodes extend their bonding with the most frequent entities. One KG unveils the correct knowledge pattern of “Latin, learn, clever, Greek”, and it elucidates that brilliant boys were taught Greek and Latin languages because most of the modern and scientific knowledge books were present in these languages and were considered the language of academia. One significant knowledge discovery shows that nodes of “Greek”, “Latin” and “dunces” have not been interlinked because dunces

never learnt Greek and Latin but they have to learn the English language. This knowledge has been accurately connected in the KG of, "dunces, learn, English", and it shows epistemological and ontological qualities. Another knowledge pattern of "Latin, learn, distinction" manifests that academic efforts in learning Latin language lead to distinctions and better rewards. Moreover, KG of "Latin, discernment" verifies textual information that Mr Welldon expresses his discernment in comprehending hidden abilities of Churchill's Latin prose in the entrance exam. Therefore, this KG is based on ontological qualities and Links tool creates the accurate KG. In figure 2, KG has been shown and it has been supported by textual evidence. Thus, 100% accurate KG has been drawn since "Accurate, up-to-date knowledge (currency) is the aim of all connectivist learning activities" (Siemens, 2006).

Hitch--- Hiking across the Sahara by G. F. Lamb

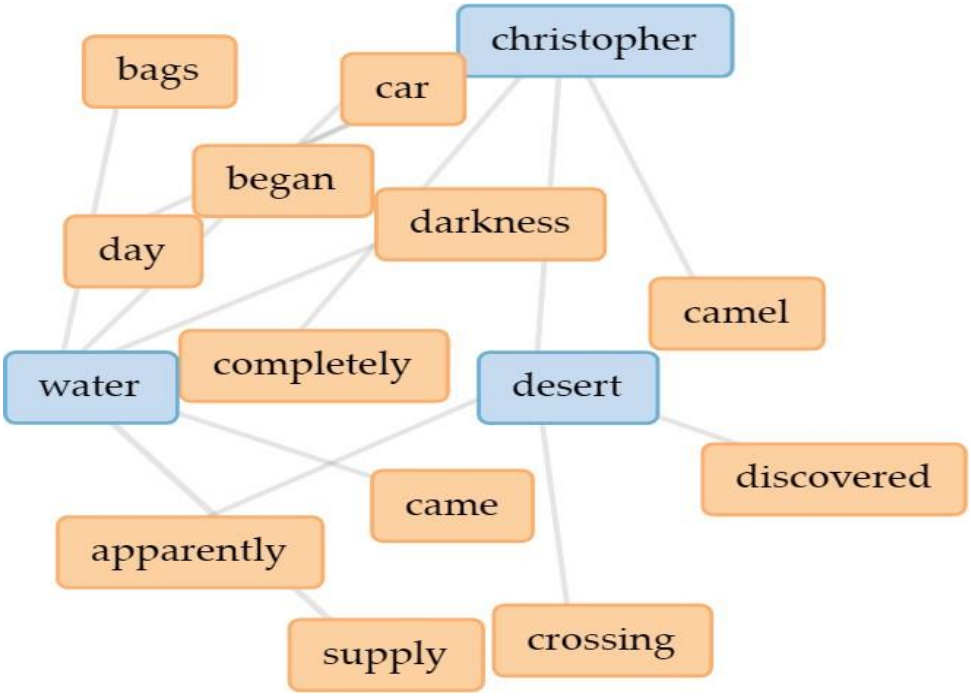

Figure 3 Hitch Hiking across the Sahara

Three blue nodes are the most frequent ones and they construct a KG of "water, Christopher, desert". "Learning is a process of connecting specialised nodes or information sources" (Siemens, 2006). It shows that Christopher, a desert hitchhiker, suffers from a shortage of water during his desert journey. He waits at In Abbangarit and is on the verge of death due to a water shortage. Then he hit upon an idea to draw water from the well with the help of a cord which was prepared from his twisted tape.

"Nurturing and maintaining connections between fields, ideas, and concepts is a core skill" (Siemens, 2006) in Knowledge graph construction. Second KG "began, Christopher, desert, crossing" verifies the fact that he begins his one desert journey after another to reach his final destination Timbuktu in the Sahara

Desert. Third KG "bag, water, supply" exposes that he stores water in animal skin bags and puts them on the camel back. Then he starts his journey from Tamanrasset to Timbuktu. Fourth KG "crossing, desert" and "car, began, day" verify another horrible incident which is narrated by a truck driver. He relates that three-car passengers try to cross the Sahara Desert but their car jams in big sand dunes. They do not carry water bags with them and one of them tries to drink oil in place of water. Eventually, their dry dead bodies have been found on scorching sand dunes.

As the accuracy of KGs is concerned, from a total of six KGS, four KGs have been proved correct and two are incomplete or incorrect. So, the proportion of their accuracy and inaccuracy is 66.66% and 33.34% respectively.

***Sir Alexander Fleming* by Patrick Pringle**

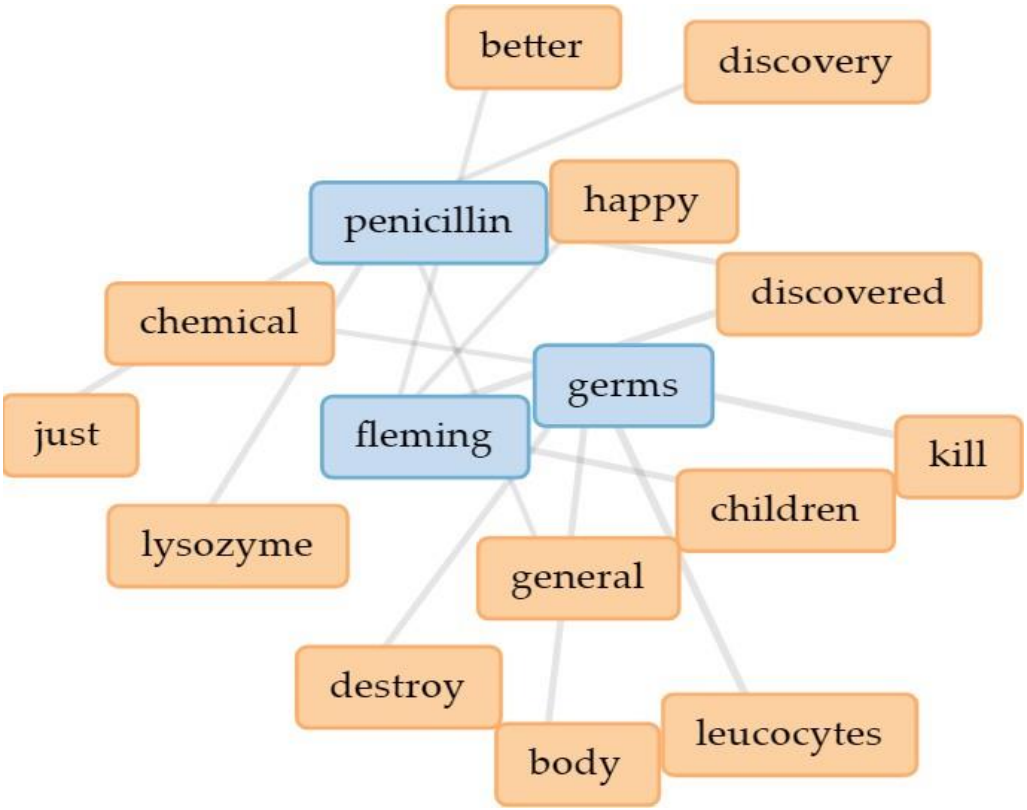

Figure 4 KG of Alexander Fleming

The blue-hued nodes "Fleming", "germs" and "penicillin" are directly linked with one another to show their interrelationship. First KG "Fleming, discovered, lysozyme" point out the first discovery of Lysozyme by Fleming, and it paves the way for other remarkable discoveries. Second KG, "kill, germs, chemical", explains that chemical kills germs. Fleming's second scientific achievement has been narrated in the third KG "discovered, penicillin, discovery" because it has been discovered by chance. Fourth KG "chemical, germs, kill" elucidates those chemicals exterminate germs as well as the patient. Consequently, their usage

causes more harm than benefits. Fifth KG "children, Fleming" highlights an event in which a man fetches his three children close to Fleming reminding them to pray for Fleming throughout their lives because his penicillin discovery has saved their lives. One knowledge discovery is that penicillin and lysozyme are same in function and both of them kill germs naturally, that's why they have been interlinked in the knowledge graph "penicillin, lysozyme". To conclude, the entity-relationship of variables (Barahate, 2012, p. 13) has been created through KGs. In this essay, five knowledge graphs have been discussed, and all of them have been verified by textual evidence. Thus, its accuracy rate is 100% and no ambiguous or incorrect KG is found. The same has been discussed in the Connectivism theory that "Accurate, up-to-date knowledge is the aim of all connectivist learning activities" (Siemens, 2006).

Louis Pasteur by Margaret Avery

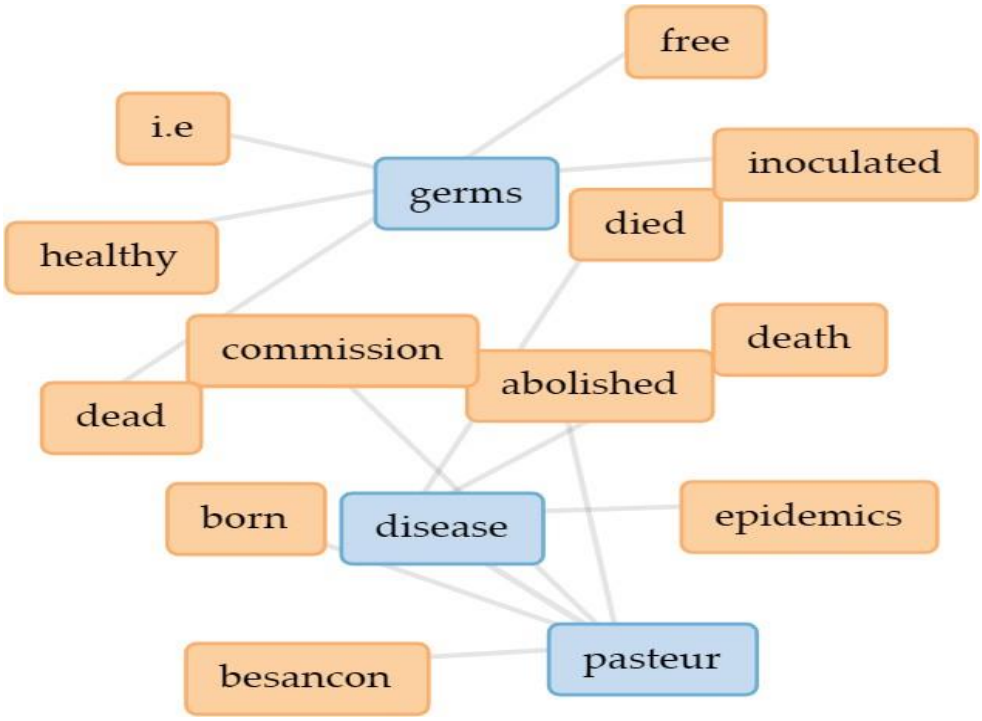

Figure 5: KG of Louis Pateur

Connectivists theory emphasises that "Learning is a process of connecting specialised nodes or information sources" (Siemens, 2006). The blue-coloured first KG "Pasteur, disease" manifests that Pasteur's seminal work is on diseases, therefore, his main contribution is the eradication of diseases from humans, animals, birds and insects. Second KG "inoculated, germs, dead" expresses his discovery of vaccination method which has used dead or dying germs as source material. Third KG "inoculated, germs, healthy" reveals the opposing themes of "disease" and "healthy" because dead germs have been inoculated to improve the health of sick fowls, animals and people. Fourth KG "Pasteur, epidemics, abolished" informs that Pasteur not only protects humans but also animals, birds and insects from lethal death. He also strengthens his economy and wins friends

by the latest brewing methods. Fifth KG of "Pasteur, commission" points out his seminal work to prove that only living beings give birth to other living beings. So, the commission falsifies the previous centuries-old notion that non-living can produce living beings and the commission declares Pasteur's victory over his opponents. Sixth KG of "Pasteur, Besancon" informs us about his admission to Besancon for graduate study and it is the very place where he met his future wife. It seems that there is a shortcoming that node "born" is not linked to any other node. Apparently, it seems an error but the expansion of KG and interactive nature of Links tool can rectify this error. This extension is possible due to the interactive nature of the tool. In this biographical essay, seven KGs have been discussed and six of them have been supported by the text; hence, its accuracy is 85.72% and one KG is incomplete and its proportion is 14.28%. As a whole, most of the KGs generated by Links tool have been proved accurate and all KGs show epistemological and ontological features which validate the tool and KGs. Thus, "Accurate, up-to-date knowledge is the aim of all connectivist learning activities" (Siemens, 2006), and this aim has been fulfilled through the current study.

Mustafa Kamal by Wilfrid F. Castle

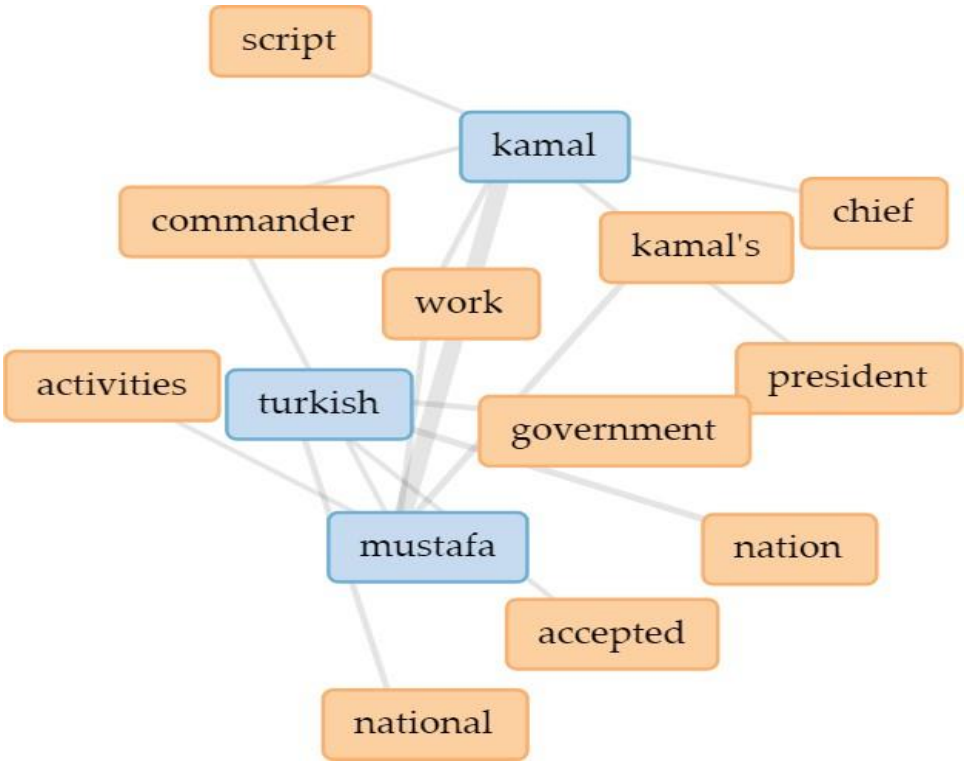

Figure 6 KG of Mustafa Kamal

"Nurturing and maintaining connections between fields, ideas, and concepts is a core skill" (Siemens, 2006) in Knowledge graph construction. The blue-coloured nodes are the most frequent; so, the first KG of "Kamal, Mustafa, Turkish" expresses that Mustafa Kamal belongs to Turkish origin, and he serves his Turkish nation by liberating them from the clutches of the allied forces. If

he was not there to unify his valiant nation to defeat the allied forces, there would not be any Turkey on the globe of the world except its small capital. The second KG of "chief, Kamal, commander" expresses his services as commander in chief of sovereign Turkey. Third KG "president, Kamal's, Kamal" expounds that the Turkish nation accepts and follows Mustafa Kamal as its president, a democratic leader and a dictator simultaneously. He was democratic because he was endeared by his all countrymen. On the other hand, he was a dictator because he snatched the throne of Mehmet. The fourth KG of "Mustafa, Kamal, script" highlights his determination and struggle to change the old and difficult script of the Ottoman Empire into Turkish script. Therefore, he leads an academic and nationalistic revolution for the sovereignty of his nation. In the discussion section of this biographical essay, four KGs have been discussed and all of them have been grounded in epistemology and ontology. Consequently, the accuracy rate of these KGs is 100%.

Measurement of Knowledge Graph Accuracy and Inaccuracy

regard to the second research question, after the construction of KGs, there is a challenge whether the generated KGs are accurate or not. In table 1, the accuracy and inaccuracy of KGs have been presented.

Table 1: Scoring for Accuracy and Inaccuracy of KGs

Essay Sr. No	Total KGs	Number of Correct KGs	Correct Percentage	Number of Incorrect KGs	Incorrect Percentage
1	5	5	100%	0	0%
2	6	4	66.66%	2	33.34%
3	5	5	100%	0	0%
4	7	6	85.7%	1	14.28%
5	4	4	100%	0	0%
	27	**24**	**88.88%**	**03**	**11%**

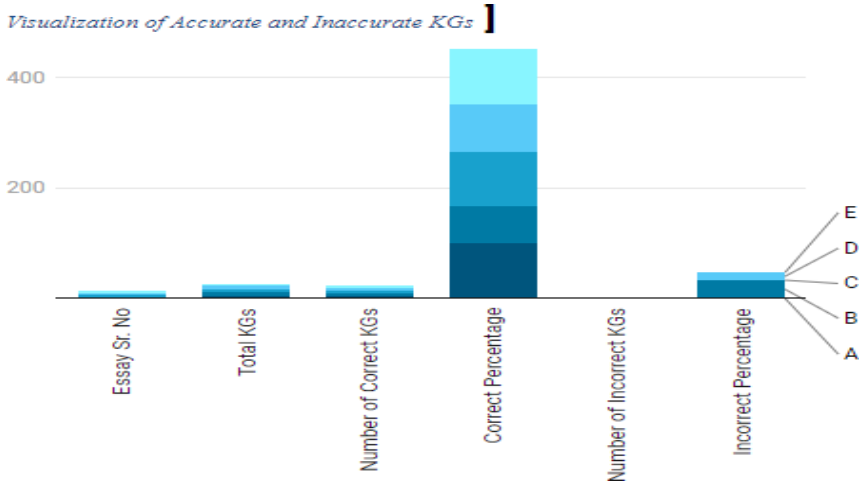

Figure 7 Comparison of Accurate and Inaccurate KGs

In Table 1, biographical essays 1, 3, 5 show 100% correct KGs while biographical essay 2, shows 66.66% accuracy and 33% inaccuracy. Essay 6 shows 85.7% accurate KGs and 14.2% inaccurate KGs. In a holistic analysis of these KGs, a total of 27 KGs have been discussed and 24 of them are accurate while three of them are inaccurate, so they need extension and disambiguation. In another way, 88.88% of KGs are accurate and 11% of KGs are inaccurate. Figure 6 also visualizes a higher percentage of correct KGs and a very less number of incorrect KGs. Such a low percentage of error is acceptable and common in machine learning processes. Table 1 and figure 7 show that Links tool in Voyant suite generates accurate KGs and they have been generated on the basis of epistemology and ontology. So, the current study proposes the use of Links tool generated KGs for distant reading, the connection of ideas and knowledge discovery for hermeneutic purposes. To conclude, "Accurate, up-to-date knowledge is the aim of all connectivist learning activities" (Siemens, 2006) and Hermeneutica Theory, so the present study conforms to its aims.

Replication of Brain Neurons as KGs

Human beings imitate natural things to make artificial things as human beings replicate brain neurons to delineate interconnected ideas in the form of knowledge graphs. As human beings interpret things with respect to different connected ideas, similarly KGs lead to a connection of linked open data, multidimensional hermeneutic interpretations and pedagogic purposes. KGs of the current study validate functions of KnowEdu system KGs (Chen, Lu, Zheng, Chen, & Yang, 2018) and support educational data mining works.

Zooming and Filtering as a Reflective Technique

"Nurturing and maintaining connections between fields, ideas, and concepts is a core skill" (Siemens, 2006) in Knowledge graph construction. The human brain does not accept the random association of one idea with every other idea, rather it zooms and filters until it reaches a logical connection, rational conclusion and in-depth meanings. Likewise, KGs as digital hermeneutic visuals give a better understanding with zooming and filtering techniques. KGs show interconnectivity among characters, settings and activities of each personality. Thus, relationship mining among different potential variables (Barahate, 2012, p. 13) has been proved in the form of data visualization of KGs.

Limitations and Future Research

This study is limited in some aspects and future researches may cover up those deficient dimensions. Firstly, the current study is an experiment on the text but a user study should be conducted to know the pros and cons of the Links tool and how does a user respond to the knowledge graph and how does a KG enhance users' hermeneutic ability. Second, the current study concentrates on one tool and comparative studies of different tools should be conducted in order to find the most productive tool for educational data mining. Lastly, the pedagogical experiment of Links tool has not been done in this study and it carries future research potential in it.

CONCLUSION

Knowledge graphs produce epistemological and ontological features and they conform to the source text. Major findings of the current study reveal that 88.88% of KGs verify textual evidence and 11% of KG nodes show no entity relationship. This minor error can be overcome by the extension of KGs and such interactivity is present in Links tool from Voyant suite. Secondly, the current study also verifies the validity and reliability of Voyant suite, and it proves the claim of tool designers that they are "knowledge bearing tools" (Rockwell & Sinclair, 2016, p. 166). Moreover, it verifies the theoretical underpinning that "Accurate, up-to-date knowledge is the aim of all connectivist learning activities" (Siemens, 2006). Thirdly, KGs develop those connections which may be ignored by human cognition. Thus, the hermeneutic capabilities of human beings enhance the extraction of knowledge patterns from big data. The major contribution of the current study is to extract knowledge graphs from small or voluminous texts quickly. Its implications are multifaceted: they can be applied for educational data mining and academic studies with distant reading. In addition, developing the association of different entities from big data has become easy, hermeneutic and time-saving. Moreover, coherence and cohesion in essays and articles should be evaluated with Links Voyant suite.

Implications for Practice and Policy

This study can be applied to the following domains:

- The educators should use knowledge graphs for teaching and learning in their classrooms. Thus, educational data mining can be promoted for learners and teachers for pedagogical purposes.
- Instructional designers should include knowledge graphs of each unit in textbooks.
- Policymakers should proclaim and highlight the need for knowledge graphs when learners come across big data and numerous books in a library on one topic.
- Administrators can improve administrative tasks by finding knowledge patterns and linking data with other themes and activities.
- Researchers can apply Links tool to develop associations with other types of data for instance social media data etc.

RECOMMENDATIONS

Future researches can be conducted to generate knowledge graphs from web data and social media pages with Links tool. Besides, KGs can be applied to the industry for finding interested target customers. Those who are reading some material, KGs will interlink them, their contact numbers and emails with the product and e-commerce companies will contact the person immediately. Secondly, the application of KGs on medical data has produced an association of health data and the discovery of knowledge patterns from big data. To conclude, implications based on the current study suggest that Links tool and its generated KGs are promising for educational data mining.

REFERENCES

Baalen, R. V. (2012). The singing narrators of fictional lies: A close and distant reading of Dutch mendacious songs (Master's thesis, Utrecht University, Utrecht, Netherland). Retrieved from https://dspace.library.uu.nl/handle/1874/254725

Barahate, S. R. (2012). Educational data mining as a trend of data mining in educational system. In IJCA Proceedings on International Conference and Workshop on Emerging Trends in Technology (ICWET 2012) icwet (9): March (pp. 11-16).

Blanco, R. & Zaragoza, H. (2010). Finding support sentences for entities. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. pp 339–346. ACM.

Bouarroudj, W., Boufaida, Z., & Bellatreche, L. (2022). Named entity disambiguation in short texts over knowledge graphs. Knowledge and Information Systems, 1-27.

Buitrago, M., & Chiappe, A. (2019). Representation of knowledge in digital educational environments: A systematic review of literature. Australasian Journal of Educational Technology, 35(4), 46-62. https://doi.org/10.14742/ajet.4041

Chen, P., Lu, Y., Zheng, V. W., Chen, X., & Yang, B. (2018). KnowEdu: A System to Construct Knowledge Graph for Education. IEEE Access, 6, 31553-31563. DOI: 10.1109/ACCESS.2018.2839607

Churchland, P. S., & Sejnowski, T. J. (1992). The Computational Brain. MIT Press.

Downes, S. (2006). Learning Networks and Connective Knowledge. Retrieved from http://itforum.coe.uga.edu/paper92/paper92.htlm

Ehrlinger, L., & Wöß, W. (2016). Towards a Definition of Knowledge Graphs. SEMANTiCS (Posters, Demos, SuCCESS), 48.

Fang, L., Sarma, A. D., Yu, C., & Bohannon, P. (2011). Rex: explaining relationships between entity pairs. Proceedings of the VLDB Endowment. 5(3):241–252. DOI: 10.14778/2078331.2078339

Goldie, J. G. S. (2016). Connectivism: A knowledge learning theory for the digital age?. Medical teacher, 38(10), 1064-1069. DOI: 10.3109/0142159X.2016.1173661

Gurney, K. (2014). An introduction to neural networks. CRC Press.

Ifenthaler, D., Gibson, D., & Dobozy, E. (2018). Informing learning design through analytics: Applying network graph analysis. Australasian Journal of Educational Technology, 34(2). DOI: https://doi.org/10.14742/ajet.3767

Kasneci, G., Suchanek, F. M., Ifrim, G., Ramanath, M., & Weikum, G. (2008, April). Naga: Searching and ranking knowledge. In Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference. pp. 953-962. IEEE.

Koc, M. (2012). Pedagogical knowledge representation through concept mapping as a study and collaboration tool in teacher education. Australasian Journal of educational technology, 28(4). DOI: https://doi.org/10.14742/ajet.833

Mao, Y. (2021). Summary and Evaluation of the Application of Knowledge Graphs in Education 2007–2020. Discrete Dynamics in Nature and Society, 2021.

McCusker, J. P., Erickson, J., Chastain, K., Rashid, S., Weerawarana, R., & McGuinness, D. (2018). What is a knowledge graph. Semantic Web J.(submitted).

Mor, Y., & Craft, B. (2012). Learning design: Reflections upon the current landscape. Research in Learning Technology, 20(Supplement), 85–94. https://doi.org/10.3402/rlt.v20i0.19196

Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). A review of relational machine learning for knowledge graphs. Proceedings of the IEEE, 104(1), 11-33. DOI: 10.1109/JPROC.2015.2483592

Page, R. (2019). Ozymandias - a biodiversity knowledge graph. Retrieved August 17, 2019, from https://ozymandias-demo.herokuapp.com/

Papamitsiou, Z. K., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. Journal of Educational Technology and Society., 17(4), 49-64.

Paulheim, H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic web, 8(3), 489-508. DOI: 10.3233/SW-160218

Paulsen, P. S., Perera, B. E., & Kaddoura, M. H. (2020). Corporate Social Responsibility and its Effect on Employee Retention Intention: A Quantitative Study of Employee Needs, Motivation Factors and Hygiene Factors.

Pirrò, G. (2015, October). Explaining and suggesting relatedness in knowledge graphs. In International Semantic Web Conference (pp. 622-639). Springer, Cham.

Plake, C., Schiemann, T., Pankalla, M., Hakenberg, J., & Leser, U. (2006). AliBaba: PubMed as a graph. Bioinformatics, 22(19), 2444-2445. https://doi.org/10.1093/bioinformatics/btl408

Rockwell, G., & Sinclair, S. (2016). Hermeneutica: Computer-assisted interpretation in the humanities. London, England: MIT Press.

Sang, S., Yang, Z., Wang, L., Liu, X., Lin, H., & Wang, J. (2018). SemaTyP: a knowledge graph based literature mining method for drug discovery. BMC bioinformatics, 19(1), 193. https://doi.org/10.1186/s12859-018-2167-5

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. Journal of educational psychology, 91(4), 690. https://doi.org/10.1037/0022-0663.91.4.690

Siemens, G. (2005). Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning. 2(1). Retrieved from http://www.itdl.org/Journal/Jan_05/article01.htm

Siemens, G. (2006). Knowing Knowledge (1st ed.). Morrisville, NC: Lulu.com.

Tweedie, F.; Singh, S.; and Holmes, D. (1996). Neural network applications in stylometry: The federalist papers. Computers and the Humanities 30(1):1–10. https://doi.org/10.1007/BF00054024

Uszkoreit, H., & Xu, F. (2013, October). From Strings to Things SAR-Graphs: A New Type of Resource for Connecting Knowledge and Language. In NLP-DBPEDIA@ ISWC.

Voskarides, N., Meij, E., Tsagkias, M., De Rijke, M., & Weerkamp, W. (2015). Learning to explain entity relationships in knowledge graphs. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (Vol. 1, pp. 564-574). DOI: 10.3115/v1/P15-1055

Waugh, S., Adams, A., & Tweedie, F. (2000). Computational stylistics using artificial neural networks. Literary and Linguistic Computing, 15(2), 187-198. https://doi.org/10.1093/llc/15.2.187

Whitelock, D., Field, D., Pulman, S., Richardson, J. T., & Van Labeke, N. (2014). Designing and testing visual representations of draft essays for higher education students. Retrieved December, 2019 from http://oro.open.ac.uk/id/eprint/41845