International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Clustering Approach for Information Retrieval Using A Quantum-Based Computation Technique

Rupam Bhagawati[1*], Thiruselvan Subramanian[2]

Submitted: 06/12/2023 **Revised**: 20/01/2024 **Accepted**: 30/01/2024

Abstract: Today's era of the internet and digitalization requires information of various forms. Information in any configuration is predominant in performing various tasks like management and retrieval. Static information is mostly present in documents and to perform retrieval, browsing, and managing this kind of information, we have many strategies available in classical form. Many classical forms can also be categorized from high level to low level. To some extent, the realm of Quantum Computing is also employed for the task and many contemporary researchers state efficient algorithms for the task. Relevant information per the user's need is the most prominent goal of an Information System. Documents play an important role in information in this regard as well as keeping uncertainty on the relevancy of correct information as per the requirement. Documents are representations of all kinds of information related to numerous fields like academia, media, law, engineering, geography, and many more. A huge collection of information representation is scattered everywhere throughout the logical world of the internet. The collection is a blender of all types of documents from different fields and semantics. Searching and sorting complete relevant documents from this gigantic blender according to information need is an onerous task to an extent. Grouping the same information in clusters brings a change to the relevancy rate of the required information. Quantum mechanics track towards Quantum information processing provides a realm for clustering of information in the form of acquaintances. Using quantum computation in the realm of quantum mechanics, an algorithm is proposed for the task which will lead to the grouping of information by considering the microscopic properties of each and every acquaintance.

Keywords: *Quantum Information Processing; Quantum Algorithm; Clustering; Quantum computation Technique; Information Retrieval.*

1. Introduction

Information representation and storage is the predominant aspect of the Information Retrieval (IR) process. Information Retrieval (IR) systems enfold the accessibility and organization of information as per user requirement. Information in any form has to be retrieved by a retrieval system as per the user's needs, escalating the importance of the system [1]. Information processing (IP) is the mechanism to perform retrieval of data which are the requirements of users in different modes and fields. As per [2] for retrieval of information, numerous models are developed, and those are analytically presenting their retrieval results by performing processing in the traditional manner in all arenas with the help of various corresponding models.

IP can be carried out classically by representing the information in the form of general computational bits 0 and 1 [3]. Classical processing is inadequate to represent the information in other forms acceptable to systems like quantum systems. In quantum systems, information can be present in any state: the polarized state of photons, the electronic state of an atom, the spin state of an atomic nucleus, or the electrodynamical state of superconducting circuits [4]. To carry out IP in quantum systems information can be present in any of the states according to their requirements for processing [5]. Processing can be carried out by representing the information in quantum bits 0 and 1 at the same time. Quantum systems can collaborate with different states at a time by the principle of superposition once information is represented in qubits [6].

1.1 Quantum Information Processing

Quantum Information Processing furnishes an entirely new prototype for computations and communications. This processing system always tends to make intelligible how quantum physics law tackles to upgrade accession, transmission, retrieval, and processing of information [7]. Information processing in the realm of quantum developed within years to be functional in many hypotheses. Different learnings are also deployed in the quantum realm to carry out many methods like classifications and clustering for better processing of information [8]. As per quantum mechanics measuring a microscopic scale of particles that can be representable in qubits is possible and in the term of information, different properties are used as the particles in the field of quantum mechanics [9]. Quantum information processing is certainly focused on the domain of fundamental research with its particular objectives and applications in computation, correlation, and data handling in the entirety

[1*]*M.TECH, Dept. of Computer Science and Engineering, Presidency University, Bangalore, Karnataka, India*
[2]*Ph.D, Dept. of Computer Science and Engineering, Presidency University, Bangalore, Karnataka, India*

of its angles [6]. Quantum systems work on the basis of two concepts whenever it has to represent qubits in different forms [10]. Superposition and Entanglement are two concepts of quantum-based systems that have a major impact on representation [10].

The superposition theorem can encode different forms of information through parallel computation. Quantum superposition is a key notion in quantum mechanics that defines a situation in which a quantum system exists concurrently in several, different states. It is frequently stated mathematically using the linear combination concept. Quantum laws allow the information to recombine in certain ways through superposition [11]. The identical task carried out in classical computers needs parallel processors which demands more memory and computation time [12].

Entanglement is another concept of quantum computers through which qubits can be entangled. Information in the form of qubits can communicate using this concept if in non-local states and their correlation is maintained [13]. The concept of entanglement is not approachable in classical methodology. Quantum information processing extracts the concept of superposition and entanglement in many ways like retrieval, browsing, relevance checking, ranking principles, etc. [14].

Quantization of Information is carried out from qubits and stretched to entanglement and superposition [14]. The basic concept of Quantization is followed by a qubit which is a quantum analog of classical computing. The detail dealing with qubits is that they can exist in superposition states which clarifies that an electron can be on different orbits even if it is of the same atom [15]. A qubit can be described as in Equation 1.

$$|\psi> = \alpha|1> + \beta|0> \quad (1)$$

in which 0,1 are amplitudes of classical states with complex numbers that are α and β which can be a normalization condition whenever |α|^2+|β|^2=1 [16]. As stated by Dirac notation measurements of state |ψ> is irreversible as it collapsed previous values observed for |0>and |1> and loss of all memory for former α and β [17]. The superposition theorem is applied to store information in the form of qubits such that qubits can be replaceable with new information whenever required [18].

Nowadays, the existence of Information in the world is at the rate of infinity. One single area or field of knowledge can have a lot of information in correspondence to a topic from the area [19]. Information about a single topic may have various forms, structures, representations, accessibility issues, sources, etc. concerning the requirement of the user [19]. Information retrieval is the study to perform these requirements of users as per their needs. There are plenty of information retrieval systems, performing classical functionality to carry out the retrieval of information in the best possible way [20].

To retrieve the most relevant information from all its existing classical retrieval systems there are many principles that are overcoming the lack of one another [21]. Researchers provide systems in terms of the most efficient and proficient for relevant information in the aspect of classical computation like the Boolean model [22], vector space model [23], probabilistic model [24], Binary independence model (BIM), BM25 [25] and many more. Classically supervised learning methodologies are also applied for retrieving information as per user-to-user needs [26]. Algorithms like K-nearest neighbor, Cosine similarities, and Divisive coefficient are deployed with the Probability ranking principle in the Binary independence model for retrieving documents that are relevant to the user's query [27].

Retrieving information in digital form elevates consequences in relevancy ratio to its extent and in this manner, quantum information processing proposed many methodologies which are scattered in different forms [28]. Quantum mechanics can compute the microscopic state of a particle and as per this concept, information as a particle can be retrieved to the extent of its existence in any part and any form.

Applying these concepts, quantum information processing brings Quantum computation with unsupervised learning methodologies and introduces the quantum clustering method for better re-organization of microscopic particles. The concept of unsupervised learning gives the ability of functionality for the representation of data in one group naturally [29]. Forming of groups based on the hidden pattern or properties of data are clustered. Problems of identification of natural clusters present in data, dimensionality reduction representation of data, and density estimation in data is a typical tasks for unsupervised learning functions in traditional computation [30]. Clustering is a process of dividing data sets into groups based on some specific properties which keep similar data points in one cluster and dissimilar data points in a different cluster. Quantum information processing is merged with unsupervised learning in the verse of machine learning to develop clustering algorithms in different arenas to cluster information as per the desired fields [31].

As in [32], Information retrieval systems can perform efficient retrieval concerning relevancy from groups of the same information sources rather than spread sources in various forms from different areas and fields. In this regard, various classical clustering methodology for information retrieval systems are developed that leads to the manifold of research on clustering for information retrieval and concluded with different clustering

algorithms viz. k-means algorithm [33], hierarchical clustering algorithm [34] of many forms. In [33] K-means algorithm is used in a classical way to convey the accuracy of clustering for information retrieval using a locally consistent factorization method. The accuracy of clustering for information in bit representation left behind many topics in the distribution of documents untouched for the clusters due to micro-level properties and behavior of documents as stated in [35].

Quantum information processing leads to a new era of retrieval and different processing idea for many aspects, where clustering can be modified through the quantization of information [36]. Quantum methodology accept information in qubits which is the smallest representation. Quantum methodology can improve the Clusterization of information as in the nanosomic state all sources of information (documents) are distributed in limited space for selection to form clusters that will bring a complete change in accuracy. Information transformation in the form of qubits is a major task to carry out to apply quantum mechanics and the type of information is also taking care of the task [37].

Clusterization in the proposed algorithm is carried out through initialization of the Schrödinger equation and the finalization of results undergoes similarity measurement, dimensionality reductions, semantic indexing in quantum technique, and coefficient variations in the common quantum realm which have a good impact on the accuracy of clusters for retrieval and ranking. All steps and states are mentioned in later sections with detailed methodology.

The clusters of sources (documents) for retrieving information further can be used to develop ranking methodology through feature selection and incompatibility checks. The complete implementation of the methodology will bring a package of novel algorithms encapsulated with quantum computing and various quantum-allied techniques for information retrieval and ranking. The algorithm with all steps from classical aspects to quantum proposed algorithm with results, comparison, and discussion is stated in detail in all the later sections of the paper.

2. Classical Clustering Method

Information industries are facing challenges of efficient and effective organization of data, as most of the data is translated into an electronic configuration which increases the volume of information repositories [38]. Different repositories, databases, and information sources are major areas to perform the concept of Information retrieval, transaction management, and data analysis.

Information in any form is important to all industries and persons who are dealing with the field of information retrieval, natural language processing, information security, and storage [39]. All these aspects require a tremendous amount of processed information in various forms to carry out different assignments related to studies, and businesses and to develop an efficient information retrieval system. Information clustering is a subset of data clustering, a technique of data mining that is a process to extract useful information from data as per the described properties. Information present in any document is static in nature and as per today's trending volume of information in documented form is growing tremendously [40-43].

Clustering is an unsupervised learning technique. Clusters are a group of objects that shares at least one common property among all present in the same group. The process that can form such groups is clustering which can produce either overlapping or disjoint partitions of all objects [44]. Static form of information is always present in documents. Documents are the biggest sources of information from various fields, concepts, and directories which can be from all corners of the universe and present in one platform i.e., the web [45]. Classification of all documents to their corresponding fields is carried out differently till now to mention their definite classes of belongings [46]. Document clustering is a major challenge in the present scenario. Apart from natural language processing, and information storage; retrieval is the most challenging task to fulfill a user's requirement of information [47]. Classification is not an up-to-the-mark process for accurate results of retrieval, considering documents for classification where classes or properties are known priori; as this can lead a document to be present in a wrong class if definite properties are not highlighted [48]. Clustering has two different types: a) Hard Clustering (Disjoint) where each document is assigned to exactly one cluster and b) Soft Clustering (Overlap) where documents are allowed to be present in multiple clusters [49].

The clustering in classical methodology can be done using different algorithms. As we can see from [50]. Soft Clustering and Hard Clustering are furthermore divided into different partitioning, hierarchical, and frequent item-set-based algorithms. In the classical world, K- means clustering algorithm takes a major place to split documents into clusters [33]. The divisive hierarchical clustering algorithm also used the bisecting K-means which is a variant of the K- means algorithm. There is another clustering algorithm in the clustering world known as the Agglomerative clustering algorithm which is more common [51]. The Unweighted Pair Group Method with Arithmetic Mean (UPGMA) which is an agglomerative clustering algorithm in a bottom-up approach of hierarchical clustering algorithm is considered as a good performer among other agglomerative algorithms [52].

The existence of quadratic time complexity of hierarchical clustering algorithm limited the use of these algorithms even if it gives better quality clustering [53]. In the classical methodology of clustering, as per [54] K-means clustering algorithms and their variants, coincidently the hierarchical algorithm takes the process of clustering to an extent and because of linear time complexity, it becomes suitable for large datasets. Apart from the advantages of this algorithm, it is thought to produce inferior clusters in classical methods. The overlapping concept in the K-means clustering algorithm is also a major problem as it is sensitive to the selection of partial partition [55].

In the account of the, IR study, information processing in a classical manner is done tremendously with different clustering algorithms like the K-means algorithm and hierarchical clustering algorithm. Classical information processing for document clustering from many decades is in the procession but still, many classical clustering algorithms are away from the issues viz. selection of appropriate features of documents and similarity measure [56], assessment of the quality of clusters [57], optimal use of memory, and considering the semantic relationship between words like synonyms [58].

3. Proposed Quantum Clustering Method

The superposition state of photons or atoms is used by quantum information processing to operate various forms of data and transfer and accumulate data in almost many forms. The quantum clustering process is a part of quantum information processing that extends its applications to different tasks of information processing in an unsupervised manner which leads to browsing, organizing, representation, and retrieving [59].

Proposing an algorithm for retrieving purposes would lead to the betterment of classical retrieving algorithms in the quantum realm. An era of research in classical retrieving techniques for enhancing the accuracy of retrieved results steered towards the concept of clustering, as well as computation for the clusters of information carried through the Quantum computation technique. The approach of Quantization of classical clustering algorithms was initiated with the idea of the quantum clustering method. The approached algorithm describes the complete procedure of clustering for large collections of data that has to be used by information retrieval systems.

The purpose of data clustering in the realm of quantum computation proposes the existence of the Schrödinger equation. The Schrödinger equation is a basic equation in quantum mechanics that defines the behavior of quantum systems such as electrons, atoms, and molecules [60]. It is a partial differential equation that connects the temporal evolution of a quantum system's wave function to its energy and is used to compute the probability distribution of a quantum system's state. To describe the energy and position of the electron of an atom in space and time, a mathematical expression is required that affects the wave nature of the electron inside the atom. The Schrodinger wave equation is a mathematical expression used to do so. A Quantum mechanical system is governed by a linear partial differential equation that depends on the wave function, which is the Schrodinger equation.

The Schrodinger equation stated that wave function ψ (r, t) correlated with a particle moving in space can be related to potential v (r, t), where r is position and t is time. The Schrodinger equation in the later phases leads with a set of energies accord to electrons in an atom. The mathematical representation of the wave function is required to find the electrons

As per the presumptions on the wave nature of almost all circumstances and entities in this universe is defined by a classical wave function [61] given in Equation 2.

$$\Delta\widehat{\varphi} - \frac{1}{c^2}\frac{\delta^2\widehat{\varphi}}{\delta t^2} = 0 \quad (2)$$

Equation 2 is a wave equation. Independent Schrödinger equation is used for the clustering of information which will draw the differences between classical theory the quantum information theory. The methodology begins with the objective that mapping the resemblance between each data point which are static documents and a fragment which is a part of the quantum system that has a certain domain around its environment. A function φ(x) from the wave equation is used to present the state of the system and the function depends on the coordinates of variable x in the ground state. The stimulate field in the domain x is given by Equation 3.

$$\varphi(x) = \sum_{i=1}^{n} e^{\frac{-(x-x_i)^2}{2\sigma^2}} \qquad (3)$$

In which n is the fragment and σ is the limit for scaling. Equation 3 is referred to as a kernel destiny estimator, usually required to estimate the probability density of a random variable. Now in this phase after the declaration, to carry out the quantum clustering process for the documents there is a requirement of Quantum Latent semantic analysis and Gaussian wave function of Schrödinger equation. As is clearly mentioned in [62] that terms (words) are playing a major role in property in a set of documents. High divergence of terms in a set of documents can analyze the presence and absence of latent semantic topics in each acquaintance as per the user query.

Quantum latent semantic analysis is used to specify the presence of a latent topic in each data by considering a subspace S where the qubits (information) are projected by Equation 4 as:

$$P_s = \sum_{k=1}^{r} |\sigma_k >< \sigma_k| \qquad (4)$$

in which $\{|\sigma_k>\}_{k=1\ldots r}$ is an orthogonal basis of subspace S and each $|\sigma_k>$ are wave functions of latent topic z_k.

Schrödinger equation is used with Gaussian wave function which is eventually used to distribute the vectors proportionally in the vector space where Hilbert space is considered for the distribution of complex quantities in the vector space. The Eigen state of the Schrödinger equation is taken for a complete explanation of the methodology from Equation 5 [63].

$$H\varphi = \left\{-\frac{\sigma^2}{2}\nabla^2 + v(x)\right\}\varphi = E\varphi \quad (5)$$

Standardized Quantum mechanical equation where σ is an independent parameter and φ(x) is the Eigen state and it is a normalized form of data. v(x) is the potential function which is represented by Equation 6.

$$v(x) = \frac{E+(\frac{\delta^2}{2})\nabla^2\varphi}{\varphi} \quad (6)$$

v(x) from Equation (6) is characterized as the potential and nominated as the physical property for biasing the clusters to be formed. Equations (5) and (6) are mechanically used to perform quantum computation for the task of clustering information. Simplifying the equation to zero (0) presented by Equation (7).

$$\left\{-\frac{\sigma^2}{2}\nabla^2 + v(x)\right\}\varphi = E\varphi = 0 \quad (7)$$

Through Equation (7) a mathematical representation of energy can be presented. Energy in the ground state remains zero always by adding a constant to the potential as shown in Equation (7) and for static acquaintances there are no dynamic changes that lead the value of energy to zero (0).

Quantum information processing inspires the development of a clustering process based on quantum principality which is proposed further with positive presumptions that bring assurance of optimality.

The proposed methodology starts with the postulation of static documents and words (terms) present in them. Terms are selected features of documents and represented as vectors. The theory of quantum Latent Semantic Analysis (qLSA) [64] is utilized to represent documents as vectors as well. Singular Vector Decomposition (SVD) [65] is an algorithm to manipulate the term-document matrix. In the formatted decomposed matrix Schrödinger equation as described above is applied for clustering the collection of documents.

Postulates applied to datasets of documents ensuing representation of terms and documents are carried out by qLSA technique and form a term-document matrix. It is qLSA to represent the matrix $TD = \{td_{ij}\}$ in which $D = \{d_i\}_{i=1\ldots n}$ and $T = \{t_j\}_{j=1\ldots m}$ are collections of documents and terms respectively. Followed by representation, Gaussian wave function φ_i of Schrödinger equation employed to prosecute document d_i in Equation 8.

$$\varphi_i(x) = \sqrt{\frac{td_x}{\sum_{j=x\ldots n}^{n} td_x}} \quad (8)$$

Documents are vectors to be represented in the vector space for quantum computation and vectors (documents) with multiple coordinates (terms) are implemented using wave function i.e., term document matrix prosecuted by Gaussian wave function as wave function matrix $\varphi \in R^{m+n}$ as in Equation 9.

$$\varphi_i(j) = \sqrt{\frac{td_{ji}}{\sum_{j=1\ldots m}^{m} td_{ji}}} \quad (9)$$

The matrix generated is decomposed by applying SVD of $\phi = U\sum V^T$ to bring down the counts of features selected. The decomposition procedure forms three matrices viz. documents as Eigenvectors (U), terms as potential V^T, and classes of other topics and sub-topics.

First r columns of matrix U that corresponds to eigenvectors formed by latent semantic space principal are selected and normalization is performed to end up the representation of documents as vectors in vector space.

Gaussian wave equation performs the distribution of term-document matrix as vectors in vector space. A scaling parameter σ pre-supposed by the vector space after the distribution of vectors, employ as a descriptor of cluster numbers after applying the algorithm.

Representation and distribution method followed by the clustering method finally which involves potential function in the Schrödinger equation shown in Equation (5). Local minima of potential energy V(x) is determined to select the cluster centers which are pre-dominant values for documents to form a single cluster. The association of documents in the same cluster is carried out using equalization between the values of the cluster center and the potential associated with each document.

The energy remains in the ground state stating the value for the same as zero (0) maintain through potential function by including constant with energy as in Equation (7). Gradient descent method [66] deploy to evaluate local minima of final potential V(x). Computed local minima of potential are cluster centers and are used by the Jaccard similarity measurement equation to measure the similarity between the center value and the document's potential value. From the collection, documents are distributed to the desired clusters as per the clusters determined by the Jaccard coefficient. The number of clusters is notifiable to settle all the documents in the desired clusters which will be formed as per the value given to the scaling parameter, σ. The clustering methodology has a property of cluster numbers, which is not recognized before the clustering

process and the value of σ fix it. The normalized value for σ is computed using Equation 10.

$$q = \frac{1}{2*\rho^2} \qquad (10)$$

3.1 Datasets

Documents with the total count of words (terms) in it is provided by a set of datasets, Reuters-21578 and TDT2 Corpus. The weight of term frequency along with document frequency is provided by the datasets.

Reuters-21578 corpus[1], a large collection of financial data mentioning economic growth and loss has 21578 documents categorized into 135 categories. To prepare the dataset, multiple categories' labels are discarded. The final MATLAB (.mat) version of Reuters-21578 has 8293 documents in 65 categories out of 5946 are training

1.http://kdd.ics.uci.edu/databases/reuters21578/reuters21578.html documents and 2347 are testing documents. The processed dataset (Reuters-21578) has a total of 18933 terms. The dataset is available for free on the website.

TDT2(Topic detection and tracking)[2] which is a collection of documents related to data newswire and broadcast news. Collection of 11201 documents categorized in 96 semantic categories collected from various sources out of which 2 are television programs, 2 are newswire and 2 radio programs after processing it left with 36771 terms. The dataset is available for free on the website.

4. Results and Discussion

The efficiency of the proposed methodology is tested on generally available datasets of economic purpose which are Reuters-21578 and TDT2. Both are standard datasets of financial data consisting of 8293 documents, 18933 words, and 10212 documents, 36771 words respectively. Similarity measurement techniques, the Jaccard measure, and Jaccard score [67] are responsible to assert the efficiency that describes the quality of the cluster. The result for the proposed algorithm is shown in the tables below, Values in Table 1 and Table 2 are of the proposed clustering algorithm in which the value of σ, is responsible for the number of clusters. The value of σ which is inversely proportional to the number of clusters compared with the clusters formed by classical K-means clustering used in the LCCF method [68], values for the same are in Table 3 and Table 4.

σ	**Clusters Formed**	**Jaccard Measure**	**Clusters Quality**
0.355(q=2.1)	9	0.2251	0.2616
0.469(q=2.07)	7	0.2192	0.2679
0.489(q=2.04)	5	0.2001	0.2889
0.519(q=2.01)	3	0.1996	0.2998

Table 1. Quality of clusters formed by the proposed algorithm for Reuters-21578.

σ	**Clusters Formed**	**Jaccard Measure**	**Clusters Quality**
0.355(q=2.1)	9	0.3281	0.7641
0.469(q=2.07)	7	0.3115	0.7233
0.489(q=2.04)	5	0.2997	0.6898
0.519(q=2.01)	3	0.2277	0.6272

Table 2. Quality of clusters formed by the proposed algorithm for TDT2.

2.https://catalog.ldc.upenn.edu/LDC2001T57.

K-value(Fixed clusters)	**Jaccard Measure**	**Cluster Quality**
9	0.3792	0.0717
7	0.3772	0.6892
5	0.3573	0.6772
3	0.2983	0.5674

Table 3. Quality of clusters formed by K-means clustering algorithm for TDT2.

K-value(Fixed clusters)	**Jaccard Measure**	**Cluster Quality**
9	0.2306	0.2416
7	0.2221	0.2566
5	0.2100	0.2762
3	0.2006	0.2872

Table 4. Quality of clusters formed by K-means clustering algorithm for Reuters-21578.

The achieved results by both algorithms are compared and represented as a comparison graph in Figure 1. Different clusters formed by the proposed quantum-based clustering algorithms with minimum and maximum σ values are represented in Figure 2 (a), (b), (c) and (d).

Fig 1. Cluster quality comparison: QCA vs. K-Means LCCF.

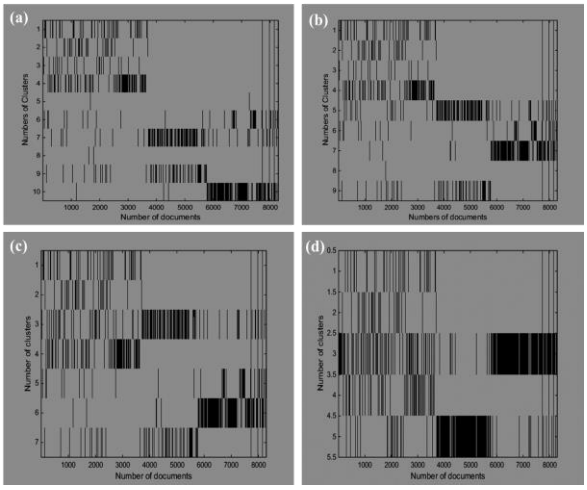

Fig 2. Clusters for Reuter Dataset with the value of (a) σ=0.355; (b) σ=0.519; Clusters for TDT2 Dataset with the value of (c) σ=0.355; (d) σ=0.519.

As a result, as discussed in the tables above, we can conclude the task of clustering by examining the cluster numbers with references to both datasets. Both datasets have different available numbers of acquaintances. As per evaluation, the Score and quality of clusters formed by the proposed algorithm for both the document sets are up to the mark than the classical K-means algorithm. We can suggest the algorithm for more relevancy tests for the information based on quality.

The resulting clusters have different documents in the form of vectors which can be further measured in the account of incompatibility by denoting the commuters between different documents which will encourage the development of novel algorithms for the relevancy of information in the realm of quantum mechanics and theory.

5. Conclusion and Future Work

The importance of the qualitative relevance algorithm is focused on in the area of information retrieval for decades, traditionally different proposal in this arena was implemented and in use for efficiency in different forms. The proposed algorithm herewith can be implemented as a novice algorithm for an information retrieval system that brings physical properties of information storage (documents) in the scenario. The algorithm can bring a nanosomic behavior of retrieval through quantum mechanics by the vectorization theorem of all terms and documents. The clusters formed by documents can behave as a base for different deserving information as per need because the group is responsible for ranking all the documents present. The algorithm is well-designed for ranking as well as for retrieval.

To improve the relevance rate of information, only clusters are not acting as base unlike for ranking of all those documents in clusters. For more favorable relevant documents from the group or cluster, feature selection for improving ranking among the documents can be implemented where documents will act as commuters for various queries and only compatible documents will be retrieved that are relevant and in a good rank as per the requirement can be improved as future scope of IR. Quantum mechanics can bring a micro-level check for the same.

References:

[1] Croft, W.B. The Importance of Interaction for Information Retrieval. in SIGIR. 2019.

[2] Singhal, A., Modern information retrieval: A brief overview. IEEE Data Eng. Bull., 2001. **24**(4): p. 35-43.

[3] Clarke, J. and F.K. Wilhelm, Superconducting quantum bits. Nature, 2008. **453**(7198): p. 1031-1042.

[4] Devoret, M.H. and R.J. Schoelkopf, Superconducting circuits for quantum information: an outlook. Science, 2013. **339**(6124): p. 1169-1174.

[5] Lachance-Quirion, D., Y. Tabuchi, A. Gloppe, K. Usami, and Y. Nakamura, Hybrid quantum systems based on magnonics. Applied Physics Express, 2019. **12**(7): p. 070101.

[6] Zoller, P., T. Beth, D. Binosi, R. Blatt, H. Briegel, D. Bruss, T. Calarco, J.I. Cirac, D. Deutsch, and J. Eisert, Quantum information processing and communication: Strategic report on current status, visions and goals for research in Europe. The European Physical Journal D-Atomic, Molecular, Optical Plasma Physics, 2005. **36**: p. 203-228.

[7] Zeilinger, A., Experiment and the foundations of quantum physics. Reviews of Modern Physics, 1999. **71**(2): p. S288.

[8] Gebhart, V., R. Santagati, A.A. Gentile, E.M. Gauger, D. Craig, N. Ares, L. Banchi, F. Marquardt, L. Pezzè, and C. Bonato, Learning quantum systems. Nature Reviews Physics, 2023. **5**(3): p. 141-156.

[9] Córcoles, A.D., A. Kandala, A. Javadi-Abhari, D.T. McClure, A.W. Cross, K. Temme, P.D. Nation, M. Steffen, and J.M. Gambetta, Challenges and opportunities of near-term quantum computing systems. arXiv preprint arXiv:.02894, 2019.

[10] Li, T. and Z.-Q. Yin, Quantum superposition, entanglement, and state teleportation of a microorganism on an electromechanical oscillator. Science Bulletin, 2016. **61**(2): p. 163-171.

[11] Bouwmeester, D. and A. Zeilinger, The physics of quantum information: basic concepts. 2000: Springer.

[12] Nielsen, M.A. and I. Chuang, Quantum computation and quantum information. 2002, American Association of Physics Teachers.

[13] Horodecki, R., P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement. Reviews of modern physics, 2009. **81**(2): p. 865.

[14] Bub, J., Quantum entanglement and information. 2001.

[15] Lehn-Schiøler, T., A. Hegde, D. Erdogmus, and J.C. Principe, Vector quantization using information theoretic concepts. Natural Computing, 2005. **4**: p. 39-51.

[16] Preskill, J., Quantum computing in the NISQ era and beyond. Quantum, 2018. **2**: p. 79.

[17] Dirac, P.A.M. A new notation for quantum mechanics. in Mathematical Proceedings of the Cambridge Philosophical Society. 1939. Cambridge University Press.

[18] Shah, N. and S. Mahajan, Document clustering: a detailed review. International Journal of Applied Information Systems, 2012. **4**(5): p. 30-38.

[19] Velden, T., K.W. Boyack, J. Gläser, R. Koopman, A. Scharnhorst, and S. Wang, Comparison of topic extraction approaches and their results. Scientometrics, 2017. **111**: p. 1169-1221.

[20] Baghel, R. and R. Dhir, A frequent concepts based document clustering algorithm. International Journal of Computer Applications, 2010. **4**(5): p. 6-12.

[21] James, D.A., The application of classical information retrieval techniques to spoken documents. 1995, Citeseer.

[22] Lashkari, A.H., F. Mahdavi, and V. Ghomi. A boolean model in information retrieval for search engines. in 2009 International Conference on Information Management and Engineering. 2009. IEEE.

[23] Shahmirzadi, O., A. Lugowski, and K. Younge. Text similarity in vector space models: a comparative study. in 2019 18th IEEE international conference on machine learning and applications (ICMLA). 2019. IEEE.

[24] Li, Y., Probabilistic models for aggregating crowdsourced annotations. 2019, University of Melbourne, Parkville, Victoria, Australia.

[25] Frinta, K. and P.P.A. Indriati, Pencarian Berita Berbahasa Indonesia Menggunakan Metode BM25. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2019. **2548**: p. 964X.

[26] Mirjalili, S., H. Faris, and I. Aljarah, Evolutionary machine learning techniques. 2019: Springer.

[27] Yang, S., G. Huang, B. Ofoghi, and J. Yearwood, Short text similarity measurement using context-aware weighted biterms. Concurrency Computation: Practice Experience, 2022. **34**(8): p. e5765.

[28] Gringoli, F., M. Schulz, J. Link, and M. Hollick. Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets. in Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization. 2019.

[29] Dervakos, E., G. Filandrianos, K. Thomas, A. Mandalios, C. Zerva, and G. Stamou. Semantic Enrichment of Pretrained Embedding Output for Unsupervised IR. in AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering. 2021.

[30] Gao, C., G. Bian, Y. Dong, X. Yuan, and H. Liu. Infrared Image Captioning Based on Unsupervised Learning and Reinforcement Learning. in 2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE). 2022. IEEE.

[31] Goswami, M. and B. Purkayastha, A Fuzzy Based Approach for Empirical Analysis of Unstructured Data. Journal of Computational Theoretical Nanoscience, 2020. **17**(9-10): p. 4375-4379.

[32] Liu, C., Y.-H. Liu, J. Liu, and R. Bierig, Search interface design and evaluation. Foundations Trends® in Information Retrieval, 2021. **15**(3-4): p. 243-416.

[33] Abualigah, L.M.Q., Feature selection and enhanced krill herd algorithm for text document clustering. 2019.

[34] Benabdellah, A.C., A. Benghabrit, and I. Bouhaddou, A survey of clustering algorithms for an industrial context. Procedia computer science, 2019. **148**: p. 291-302.

[35] Li, L., Q. Lin, and Z. Ming, A survey of artificial immune algorithms for multi-objective optimization. Neurocomputing, 2022. **489**: p. 211-229.

[36] Huleihel, W., A. Mazumdar, M. Médard, and S. Pal, Same-cluster querying for overlapping clusters. Advances in Neural Information Processing Systems, 2019. **32**.

[37] Bhagawati, R. Clusters Analyzer Algorithm for Informative Acquaintances-Quantum Clustering Algorithm. in 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). 2020. IEEE.

[38] Hu, X., L. Leydesdorff, and R. Rousseau, Exponential growth in the number of items in the WoS. ISSI Newsletter, 2020. **16**(2): p. 32-38.

[39] Verma, R.M. and S. Srinivasagopalan. Clustering for security challenges. in Proceedings of the ACM International Workshop on Security and Privacy Analytics. 2019.

[40] Afzali, M. and S. Kumar. Text document clustering: issues and challenges. in 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon). 2019. IEEE.

[41] Sirichanya, C. and K. Kraisak, Semantic data mining in the information age: A systematic review. International Journal of Intelligent Systems, 2021. **36**(8): p. 3880-3916.

[42] Zheng, Z., X. Li, M. Tang, F. Xie, and M.R. Lyu, Web service QoS prediction via collaborative filtering: A survey. IEEE Transactions on Services Computing, 2020. **15**(4): p. 2455-2472.

[43] Bhagawati, R., S.R. Laskar, and B. Swain. Documents clustering using quantum clustering algorithm. in 2016 International Conference on Microelectronics, Computing and Communications (MicroCom). 2016. IEEE.

[44] Ammar, H.A. and R. Adve. Power delay profile in coordinated distributed networks: User-centric v/s disjoint clustering. in 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP). 2019. IEEE.

[45] Zamani, H., S. Dumais, N. Craswell, P. Bennett, and G. Lueck. Generating clarifying questions for information retrieval. in Proceedings of the web conference 2020. 2020.

[46] Eminagaoglu, M., A new similarity measure for vector space models in text classification and information retrieval. Journal of Information Science, 2022. **48**(4): p. 463-476.

[47] Meng, Y., Y. Zhang, J. Huang, Y. Zhang, and J. Han. Topic discovery via latent space clustering of pretrained language model representations. in Proceedings of the ACM Web Conference 2022. 2022.

[48] Barshandeh, S., R. Dana, and P. Eskandarian, A learning automata-based hybrid MPA and JS algorithm for numerical optimization problems and its application on data clustering. Knowledge-Based Systems, 2022. **236**: p. 107682.

[49] Raut, A. and G. Bamnote, Soft clustering: An overview. IJCCT, 2010. **1**(2-4 SPEC. ISSUE): p. 370-372.

[50] Banerjee, A., S. Merugu, I.S. Dhillon, J. Ghosh, and J. Lafferty, Clustering with Bregman divergences. Journal of machine learning research, 2005. **6**(10).

[51] Murtagh, F., Hierarchical Clustering. 2011.

[52] Li, T., A. Rezaeipanah, and E.M.T. El Din, An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement. Journal of King Saud University-Computer Information Sciences, 2022. **34**(6): p. 3828-3842.

[53] Ambroise, C., A. Dehman, P. Neuvial, G. Rigaill, and N. Vialaneix, Adjacency-constrained hierarchical clustering of a band similarity matrix with application to genomics. Algorithms for Molecular Biology, 2019. **14**(1): p. 22.

[54] Ahmed, M., R. Seraj, and S.M.S. Islam, The k-means algorithm: A comprehensive survey and performance evaluation. Electronics, 2020. **9**(8): p. 1295.

[55] Khan, S.U., A.J. Awan, and G. Vall-Llosera, K-means clustering on noisy intermediate scale quantum computers. arXiv preprint arXiv:.12183, 2019.

[56] Pistoia, M., S.F. Ahmad, A. Ajagekar, A. Buts, S. Chakrabarti, D. Herman, S. Hu, A. Jena, P. Minssen, and P. Niroula. Quantum Machine Learning for Finance ICCAD Special Session Paper. in 2021

IEEE/ACM International Conference On Computer Aided Design (ICCAD). 2021. IEEE.

[57] Arthur, D. and P. Date, Balanced k-means clustering on an adiabatic quantum computer. Quantum Information Processing, 2021. **20**: p. 1-30.

[58] Parekh, R., A. Ricciardi, A. Darwish, and S. DiAdamo. Quantum algorithms and simulation for parallel and distributed quantum computing. in 2021 IEEE/ACM Second International Workshop on Quantum Computing Software (QCS). 2021. IEEE.

[59] Steinbach, M., G. Karypis, and V. Kumar, A comparison of document clustering techniques. 2000.

[60] Tsutsumi, Y., Schrodinger equation. Funkcialaj Ekvacioj, 1987. **30**: p. 115-125.

[61] Horn, D. and A. Gottlieb, The method of quantum clustering. Advances in neural information processing systems, 2001. **14**.

[62] Amati, G. and F. Crestani, Advances in Information Retrieval Theory: Third International Conference, ICTIR 2011, Bertinoro, Italy, September 12-14, 2011, Proceedings. Vol. 6931. 2011: Springer Science & Business Media.

[63] Bernstein, D.H., E. Giladi, and K.R. Jones, Eigenstates of the gravitational Schrödinger equation. Modern Physics Letters A, 1998. **13**(29): p. 2327-2336.

[64] Ginter, F., H. Suominen, S. Pyysalo, and T. Salakoski, Combining hidden Markov models and latent semantic analysis for topic segmentation and labeling: Method and clinical application. International journal of medical informatics, 2009. **78**(12): p. e1-e6.

[65] Patil, R., Noise reduction using wavelet transform and singular vector decomposition. Procedia Computer Science, 2015. **54**: p. 849-853.

[66] Amari, S.-i., Backpropagation and stochastic gradient descent method. Neurocomputing, 1993. **5**(4-5): p. 185-196.

[67] Jain, A., A. Jain, N. Chauhan, V. Singh, and N. Thakur, Information retrieval using cosine and jaccard similarity measures in vector space model. Int. J. Comput. Appl, 2017. **164**(6): p. 28-30.

[68] Cai, D., X. He, and J. Han, Locally consistent concept factorization for document clustering. IEEE Transactions on Knowledge Data Engineering, 2010. **23**(6): p. 902-913.