

# Common Data Storage System

AnjuG S,ArathyA S,Merlin Babu,PriyaR S,SwathyJ  
Computer Science and Engineering  
College Of Engineering, Perumon  
Kerala, India

Jooby E  
Computer Science and Engineering  
College Of Engineering, Perumon  
Kerala, India

**Abstract**— This paper discuss about a Universal Framework for Knowledge Representation (UfK) that can represent and process diverse kinds of data types and formats in a single, common data storage system (CDSS). The data types handled by the CDSS can be text documents, images, audio files and/or video files. The manipulation and integration of heterogeneous data from different sources in to a meaningful common representation is a major challenge for existing systems, because different data types require different processing algorithms. The processing of each data type is complicated further by the huge variety in formats used to represent it, such as in the example of the image data type, which can be represented by BMP, PNG, JPG, TIFF and other such image formats. Another aspect of this problem is the difficulty in storing vast quantities of such data in widely dispersed data store (causes high latency) or in large centralized data servers (very high costs). The BigData paradigm solutions are that related data be stored together in cheap distributed parallel data stores. UfK provides a highly flexible data representation system based on UfKClasses that encapsulate heterogeneous data specific to that class along with extracted patterns from the data and the algorithm hints used to process that data.

**Keywords**-Bigdata;UniversalFramework;Knowledge representation;Common Data Storage System;

## I. INTRODUCTION

In this paper we discuss about a CDSS which is a data store that can represent and process diverse kinds of data types and formats in a single CDSS .Data is information processed or stored by a computer. Information can be text, image, audio and video. Different computer based formats needs to be processed in its own special way. So the processing of data becomes very difficult. CDSS helps to overcome this problem.

For example, consider the case of a large object database like AADHAR. It contains details of a person like image, text data, biometric details etc. These various format of data are stored in its own database in a horizontal manner. Thus the accessing of data becomes very difficult. These problems can be solved by the development of a CDSS in which data are stored in vertical manner. This system can solve various problems of bigdata by making use of Universal Framework for Knowledge Representation [1].

## II. ABOUT THE SYSTEM

There are no existing systems like the UfK. So this is a newly designed data storage system in which different types of data can be stored and easily retrieved .Here we used some basic concepts of the following papers.

FRONTIERS IN MASSIVE DATA ANALYSIS [2] deals with the concept of massive data analysis. It reduces statistical complexity using data representation. Two basic approaches used are dimensionality reduction and clustering. In dimensionality reduction it reexpresses the data in the form of high dimensional vectors. Sampling methods and parallel algorithms are used in these methods. Clustering is a simplified method in which data's are divided into small number of clusters. The advantage of this concept are it will reduce runtime, this promotes multi-stage approach, it convert data into more compact form, it require less storage place. The main disadvantage is that the clustering of data leads to the loss of fine details.

Next concept we used is ANALYSIS OF BIG DATA ISSUES USING SP THEORY OF INTELLIGENC [3] by Ms. Vijayashanthi.R, Mrs. N.Shunmuga Karpagam. It deals with the collection of bigdata which is difficult to process. In this a SP system which has a potential for UfK representation has been used. This deals with the problem of variety in bigdata. It also solves the velocity, volume and veracity (finding errors) of bigdata. It gives importance to the area where good results are found. The integration of memory and process takes place here.

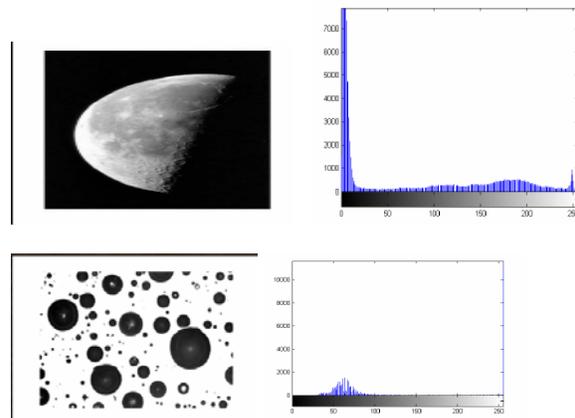


Figure 1. Some Images and their Histogram

Next concept is COLOR HISTOGRAM BASED IMAGE RETRIEVAL [8] by Chesti Altaff Hussain, Dr. D. Venkata Rao, T. Praveen in 2013 deals with Color is the essential feature for searching. In this first extract the histogram and Euclidean distance formula has been used for the comparison.

RGB color combination is used and it is calculated by measuring intensities of each color because images are represented as pixel values. It is applicable to black and white images. The advantage is it is accurate, efficient (90%) and easy. The main disadvantage is the shape and texture information are discarded, it is applicable to small database and it is not incorporate with spatial characteristics.

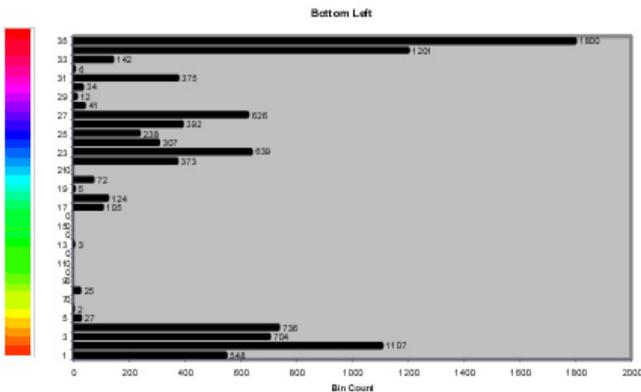


Figure 2. Color Histogram of above Picture

Next concept is COMPARING IMAGES USING JOINT HISTOGRAM [9] by Greg Pass, Ramin Zabih. The histogram details used in the project is to search for matching images and retrieving the details of that data. This paper gives information about the joint histogram, how it differs from other color histograms. The paper specifies that color histogram is widely used effectively. The use of color histogram gives information about histogram distance for each pixel in the image. In sometimes different images give matching information and can have same color histograms. The advantage of this system is it is applicable to large database.

The Universal Framework for Knowledge Representation (UFK) can represent and process diverse kinds of data types and formats in a single, common data storage system (CDSS). The data types handled by the CDSS can be text documents,

images, audio files and/or video files. The manipulation and integration of heterogeneous data [5] from different sources in to a meaningful common representation is a major challenge for existing systems, because different data types require different processing algorithms. The processing of each data type is complicated further by the huge variety in formats used to represent it, such as in the example of the image data type, which can be represented by BMP, PNG, JPG, TIFF and other such image formats. Another aspect of this problem is the difficulty in storing vast quantities of such data in widely dispersed data store (causes high latency) or in large centralized data servers (very high costs). The BigData paradigm solution is that related data be stored together in cheap distributed parallel data stores. UFK provides a highly flexible data representation system based on UfkClasses that encapsulate heterogeneous data specific to that class along with extracted patterns from the data and the algorithm hints used to process that data.

The UFK framework presents an elegant solution by providing a CDSS that can store these UFKclass in tree hierarchies, making it easier for applications developed based on the UFK Framework to leverage this paradigm to less processing costs and increased stability [4]. Since, in the UFK system, knowledge is represented as patterns for improving the processing speed of data like searching and retrieval from the CDSS. This has wide applications in fields such as image processing, search engines, data mining, large object databases (like e-governance, AADHAR) and Machine learning.

In Image processing [6], the storage and retrieval of patterns associated with an image plays a very important part in analytical algorithms, the UFK framework allows the native storage and format processing of various image types, thereby freeing the researcher. As search and match algorithms are an integral part of the UFK, Search engines that can be used to pattern match on multi-data keys (contains, for example: both an image and associated textual information). The UFK format CDSS is particularly suited for vertical partitioning allowing different subsets of UFKclass to be store in different locations, which allows for increased speed for search engines. This feature also makes it suitable for implementing large object databases and data mining operations by reducing cost of processing. The flexible structural relationship [7] between data and meta-data in UFK format makes it suitable for implementing neural network based algorithms in the inherent ecosystem of this framework.

The Proposed system description is divided in to three processes, Encoding, Decoding and Searching. Encoding is concerned with the conversion of various kinds of input data into the internal UFK representation. In searching the user will give related keywords or data or partial data constructed as an UFKclass. Using this, the system searches the UFK CDSS. Decoding deals with the encoded file, this is retrieved from the database.

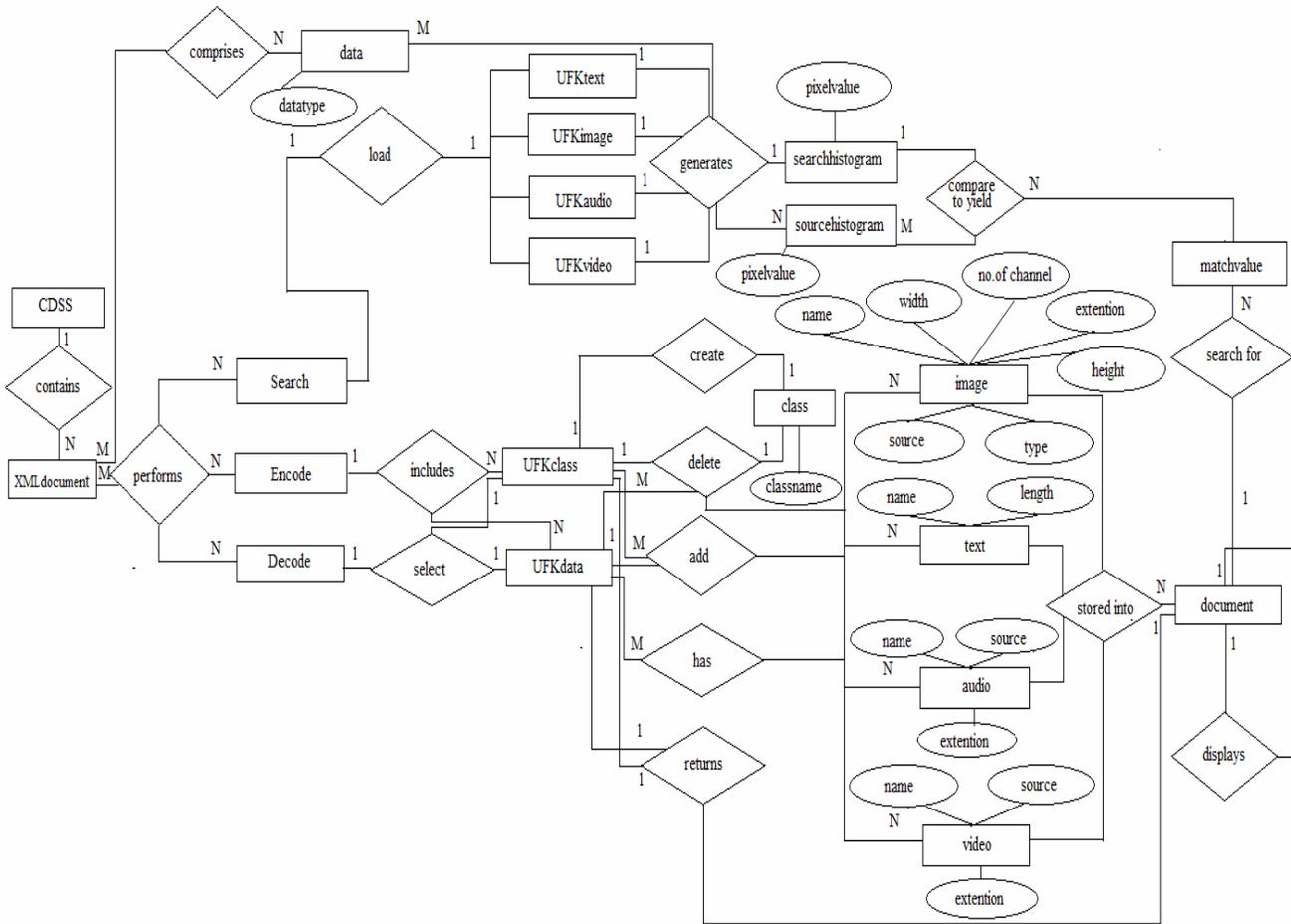


Figure 3. System Design

### III. FUTURE ENHANCEMENT

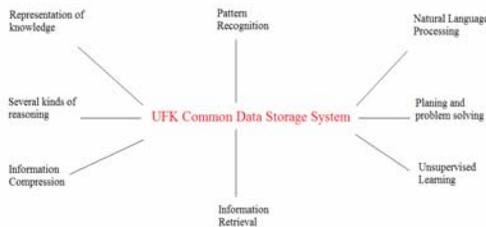


Figure 4. Area of application

The UFK system can perform several kinds of reasoning including deductive, abductive, non-monotonic, and monotonic and Bayesian reasoning. It can also achieve pattern

recognition at multiple level of abstraction. The UFK system has strength in natural language processing area that is it can generate and respond to instructions in natural language. It also finds applications in the field of planning and problem solving. It facilitates information retrieval and has the potential to create the facilities of a query language like SQL.

### IV. CONCLUSION

The main aim of this concept is a universal framework where diverse kinds of data can be stored in a common data storage system. Here the data storage system can store all types of data. This is the main advantage of the system. By the development of this system the processing like searching and retrieving all types of data is very effectively done. This data storage system can be used to many applications and projects as effective data storage. Simplify and integrate concepts across mainstream computing. UFK helping to reduce the problem of variety in big data. There are great diversity in formalism and formats for knowledge, and how they are

processed. The UFK system is likely to yield direct benefits in

the storage, management, and transmission of big data.

## REFERENCES

- [1] J. G. Wolff, "Big Data and the SP theory of intelligence," pp. 'Menai Bridge, U.K.: CognitionResearch.org, vol. 2, 301-315, 2014
- [2] National Research Council, *Frontiers in Massive Data Analysis*. Washington, DC, USA: National Academies Press, 2013.
- [3] R. Vijayashanthi, N. Shunmuga Karpagam, "Analysis of Big Data Issues Using SP Theory of Intelligence", *International Journal on Applications in Science, Engineering & Technology Volume.1, Issue.2, 2015, pp.11-14*
- [4] J. E. Kelly and S. Hamm, *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. New York, NY, USA: Columbia Univ. Press, 2013.
- [5] J. G. Wolff, "The SP theory of intelligence: An overview," *Information*, vol. 4, no. 3, pp. 283341, 2013.
- [6] J. G. Wolff, *Unifying Computing and Cognition: The SP Theory and Its Applications*. Menai Bridge, U.K.: CognitionResearch.org, 2006.
- [7] J. G. Wolff, "The SP theory of intelligence: Benets and applications", *Information*, vol. 5, no. 1, pp. 127, 2014.
- [8] Chesti Altaff Hussain, Dr. D. Venkata Rao, T. Praveen, "Color Histogram Based Image Retrieval Hussain, et al., *International Journal of Advanced Engineering Technology* .
- [9] Greg Pass, Ramin Zabih, "Comparing Images Using Joint Histograms", Computer Science Department, Cornell University, Ithaca, NY 14853 fgregpass, rdzg@cs.cornell.edu ,607-255-8413

## AUTHORS PROFILE

Anju G S  
U G scholar  
Computerscience and engineering  
College of engineering, perumon  
Kerala,India.

Arathy A S  
U G scholar  
Computerscience and engineering  
College of engineering, perumon  
Kerala,India.

Merlin Babu  
U G scholar  
Computerscience and engineering  
College of engineering, perumon  
Kerala,India.

Priya R S  
U G scholar  
Computerscience and engineering  
College of engineering, perumon  
Kerala,India

Swathy J  
U G scholar  
Computer Science and Engineering  
College of engineering, perumon  
Kerala,India.

Jooby E  
Assistant Professor in CSE  
Computer Science and Engineering  
College of engineering, perumon  
Kerala,India.