Universal Framework for Knowledge Representation


U G Scholars, Computer Science & Engineering, College Of Engineering, Perumon, Kerala, India

Assistant Professor in CSE, Computer Science & Engineering, College of Engineering, Perumon, Kerala, India

Assistant Professor in IT, Information Technology, College of Engineering, Perumon, Kerala, India

Abstract: This paper discusses 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 into 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 stores (causes high latency) or in large centralized data servers (very high costs). The Big Data paradigm solutions is that related data be stored together in cheap distributed parallel data stores. UFK provides a highly flexible data representation system based on Ufk Classes that encapsulate heterogeneous data specific to that class along with extracted patterns from the data and the algorithm hints used to process that data.

1. Introduction

In this project we are developing 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 need 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 big data.

2. Big Data

Big data [2] is the collection of large amount of data. These data can be text, image, audio and video. Due to large size of data it becomes very difficult to process. The difficulty lies in their transmission, storage, searching, capture, analysis and sharing. In addition to that variation in the amount of data stored in different areas is also a problem. UFK system can solve these problems.

3. Existing System

The manipulation and integration of heterogeneous data [3] from different sources in to a meaningful common representation is a major challenge for existing systems, because different data types require different processing algorithms. Also each kind of data have different computer based format and processed its own special way. There is no existing system to represent and process diverse kinds of data types and formats in a single, common data store system.

4. Proposed System

The proposed system named “Universal Framework for Knowledge Representation is a system which allows representation and processing of 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. 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 UfkClasses 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. In Ufk all data are stored as XML tags. Firstly we create an XML document and create suitable UfkClasses within it. We can add text documents, images, audio and video to the corresponding classes and also have the option of deleting them from CDSS. This system improves the processing speed of data like searching and retrieval from the CDSS.

In Image processing[4], the storage and retrieval of patterns associated with an image plays a very important part in analytical algorithms, the UFK framework[1] 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).

5. Module Description

The proposed system is divided in to three modules

5.1. Encoding

This module is concerned with the conversion of various kinds of input data into the internal UFK representation. There are various sub modules are for each data type processed by this system. The processed output is stored in an XML format file as it most closely resembles the tree structure of the UFK framework. Depending on data type, a pattern extraction algorithm is used and this information is added to the stored data.

5.2. Searching

The user will give related keywords or data or partial data constructed as an UFKclass. Using this, the system searches the UFK CDSS. It then outputs the match probability for each UFKclass in the CDSS with the queried class.

5.3. Decoding

The encoded file is retrieved from the database. Then consistency checking is performed. After this the in-memory file is written to physical UFK file. The original data is reconstructed from the UFK file.

6. System Implementation

It is the process of converting a new system design into operation. This system is implemented with different algorithms like:

6.1. Processing algorithm

In processing different data are stored as histogram. Histogram is the graphical representation of data. Images search can be done using histogram [3].

6.2. Comparison algorithm

This algorithm is used to compare histograms[4]. It give the matching percentage. The different algorithms are

6.2.1 Chi-square

\[ d(H_1,H_2) = \sum \left( \left( \frac{H_1(i) - H_2(i)}{H_1(i)} \right)^2 \right) \]

Let \( H_1 \) and \( H_2 \) be the histograms of two different data.

6.2.2 Intersection

\[ d(H_1,H_2) = \sum \min (H_1(i),H_2(i)) \]

6.3. Coding and Decoding algorithm

Built in java algorithms are used to perform coding and decoding.

7. Figures

![Figure 7.1 User access](image_url)
From figure 7.1 it is very clear about the system. The user can search related pattern format files. The files or data of different formats are stored in a common database. After searching, the user can view the full details about the matching data that is used for searching. The user can get all the images, videos, audios etc.

Figure 7.2 give information about how the data are represented in UFK. For example consider the case of storing a text data and image together in the database. For this first convert the data into UFK and store in XML document.

Figure 7.3 is a representation of the decoding part. Here XML file is converted to UFK format and then it is retrieved as text or image as needed by the user. Figure 7.4 shows how the searching is performed. In searching user give related information and the result of the search is the search list from the UFK. Then it is decoded to the corresponding pattern. The user gets the information like text, image etc stored in UFK related to the searching topic. It is the main advantage of this new data storage.

Figure 7.5 gives the information about the different modules used in the system. The common data storage system contains all types of data in XML document. This XML document can perform encode, decode and search operations. In encoding, the data can be encoded with class and data. The encoding of data can be done by creating a class with a class name. The system can also perform the delete operation by using the ufk class. Each class can add data. Data can be of different types such as text, image, audio and video. Each image has width, extension, name and height. All types of data are stored in a document. Decoding can be done with both class and data. The decoded data are retrieved from the document. In case of searching first load the data, the data can be text, image, audio and video. The data in document are compared with the data loaded. For this first convert the data into histogram. Both source histogram and search histogram are generated. Then the two histogram generated are compared and display the matching information based on algorithms.

8. Future Enhancement

The UFK system can perform several kinds of reasoning including deductive, abductive, non-monotonic, and monotonic 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.

9. Acknowledgement

Though words cannot express our gratitude completely, We would like to thank ever one who had speed as a pillar of strength and support for us. Above all we concede the surviving presence and flourishing refinement of the ALMIGHTY GOD for concealed hand yet substantial supervision all through the project. We would like to thank our respected Principal Dr Z A Zoya and our respected Head of the Department Ms. Deepa for giving us the opportunity to do this project. We also like to express our teachers Ms. Jooby E and Ms. Deepa K Daniel for guiding us all through our work. We sincerely thank all our staffs of the Department Of CSE for their assistance and timely advice. Last but not the least; we would like to thank our parents and friends for their sincere support…

10. References


