



# CREDIT CARD FRAUD TRANSACTION DETECTION USING MACHINE LEARNING

R.VENKADESH M.E<sup>1</sup>, ARAVINDH.D<sup>2</sup>, BOOBALAN.R<sup>3</sup>, PRAVEEN KUMAR.V<sup>4</sup>, CHANDRU.S<sup>5</sup>

1ASSISTANT PROFESSOR, 2,3,4 UG STUDENTS

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
MAHENDRA ENGINEERING COLLEGE, TAMILNADU, INDIA

## INTRODUCTION

### 1.1 INTRODUCTION

In recent years, many online social media websites allow users to both upload multimedia data and annotate the content with tags. The social tagging is foreseen as a method to bridge the semantic gap in image analysis. Tag-based search, which returns images annotated with a specific query tag is an important way of searching or browsing images on social dataset. This image search method, to some extent, has achieved some success on exploiting the associated tags for indexing and searching large-scale web images compared with text-based image search and content-based image search. Despite the success of social tagging, however, tags contributed by common users are known to be ambiguous, limited in terms of completeness, and overly personalized. This is not surprising because of the uncontrolled nature of social tagging and the diversity of knowledge and cultural background of its users. Although the relevance of a tag given the visual content can be subjective for a specific user, an objective criterion is desirable for general-purpose search and visual content understanding. We consider a tag relevant to an image if the tag accurately describes objective aspects of the visual content, or in other words, users with common knowledge relate the tag to the visual content easily and consistently.

Generally speaking, tag-based image search is more commonly used in social media than content based image retrieval and context-and-content based image retrieval. In recent years, the re-ranking problem in the tag-based image retrieval has gained researchers' wide attention. Starting from this intuition and above analysis, it proposes a social re-ranking algorithm which user information is firstly introduced into the traditional ranking method considering the semantics, social clues and visual information of images.

The contributions of this project can be described as follows:

- 1) A tag-based image search approach with social re-ranking. We systematically fuse the visual information, social user's information and image view times to boost the diversity performance of the search result.
- 2) Inter user ranking is applied to rank users images according to query given. With this ranking the good tradeoff between the diversity and relevance performance which also effectively eliminate the similar images from the same user in a ranked result.

3) Title and time stamp ranking in which the desired output will get on the basis of title information and the recent time stamp which enhance the diversity performance of image ranking system.

4) The view of an image in social media community is an important feature which indicates the click count of this image. The number of click count has been utilized to improve the relevance performance of the image retrieval results.

We take Flickr as an example to study the characteristics of social tagging. Flickr is one of the earliest and most popular social media sharing web sites and it has been intensively studied in recent years, especially on tagging characteristic, tag recommendation, etc. A recent study in reveals that users do annotate their photos with the motivation to make them better accessible to the general public. However, the tags provided by Flickr users are highly noisy and there are only around 50% tags actually related to the image. Fig. 1 illustrates an exemplary image from Flickr and its tags. From the figure we can see that only “sky” and “cloud” correctly describe the content of the given image, and the other tags are imprecise(e.g., dog, girl, etc.) or subjective1(e.g., family, city, etc.). Meanwhile, several other tags that can be useful, such as “tree” and “grass”, have not been provided. The imprecise and incomplete tagging characteristics have significantly limited the access of social media.

The imprecise tags will introduce false positives into user’s search result and incomplete tags will make the actually related images inaccessible. Therefore, it would be advantageous if a dedicated approach can be developed to improve the tags associated with social images such that they can better describe the content of the images.

**Keywords : Detection , Credit card & Social Media**

## LITERATURE SURVEY

### 1) EVALUATING THE EFFECTIVENESS OF PERSONALIZED WEB SEARCH

We present a large-scale evaluation framework for personalized search based on query logs and then evaluate five personalized search algorithms (including two click-based ones and three topical-interest-based ones) using 12-day query logs of Windows Live Search. By analyzing the results, we reveal that personalized Web search does not work equally well under various situations. It represents a significant improvement over generic Web search for some queries, while it has little effect and even harms query performance under some situations. Experimental results show that using a personalization algorithm for queries selected by our prediction model is better than using it simply for all queries

### 2) PERSONALIZED WEB SEARCH FOR IMPROVING RETRIEVAL EFFECTIVENESS

The user profiles are then used to improve retrieval effectiveness in Web search. A user profile and a general profile are learned from the user’s search history and a category hierachy, respectively. These two profiles are combined to map a user query into a set of categories which represent the user’s search intention and serve as a context to disambiguate the words in the user’s query. Web search is conducted based on both the user query and the set of categories. Several profile learning and category mapping algorithms and a fusion algorithm are provided and evaluated. Experimental results indicate that our technique to personalize Web search is both effective and efficient.

## ALGORITHMS

- mapping algorithms
- merging algorithms.

### 3) SCALABLE RETRIEVAL AND MINING WITH OPTIMAL PEER-TOPEER CONFIGURATION

The scalable peer-to-peer network to perform the content-based image retrieval and mining, i.e., P2P-CBIRM. The decentralized unstructured P2P model with certain overheads, i.e., peer clustering and update procedures, is adopted to compromise with the structured one while still reserving flexible routing control when peers join/leave or network fails. The peer CBIRM engine is designed to utilize multi-instance query with multi-feature types to effectively reduce network traffic while maintaining high retrieval accuracy. Experiments show that the query efficiency of the scalable retrieval approach is better than previous methods, i.e., firework query model and breadth-first search. It provides a scalable knowledge discovery platform for efficient image data mining applications.

#### ALGORITHM

- Retrieval algorithms.
- Mining algorithms.

#### SYSTEM ANALYSIS

##### 3.1 EXISTING SYSTEM

Thus, the tag-based image search can be easily accomplished by using the tags as query terms. However, the weakly relevant tags, noisy tags and duplicated information make the search result unsatisfactory. Most of the literatures regarding the re-ranking of the tag-based image retrieval focus on tag processing, image relevance ranking and diversity enhancement of the retrieval results.

##### Tag Processing Strategy

It has been long acknowledged that tag ranking and refinement play an important role in the re-ranking of tag-based image retrieval, for they lay a firm foundation on the development of re-ranking in tag based image retrieval (TBIR).

##### Relevance Ranking Approach

To directly rank the raw photos without undergoing any intermediate tag processing, Liu et al. utilized an optimization framework to automatically rank images based on their relevance to a given tag.

##### 3.1.1 LIMITATION OF EXISTING SYSTEM

- Most of the existing approaches highly rely on the visual and semantic information, and thus ignore the social clues such as user and view information.
- User information is always exploited to do the target advertisement, travel recommendation and user interaction based image re-ranking.
- However, user information is seldom used in retrieval work.
- 

##### 3.2 PROPOSED SYSTEM

The social re-ranking algorithm which user information is firstly introduced into the traditional ranking method considering the semantics, social clues and visual information of images.

- 1) A tag-based image search approach with social re-ranking. The systematically fuse the visual information, social user's information and image view times to boost the diversity performance of the search result.

2) The inter-user re-ranking method and intra-user re-ranking method to achieve a good trade-off between the diversity and relevance performance. These methods not only reserve the relevant images, but also effectively eliminate the similar images from the same user in the ranked results.

3) In the intra-user re-ranking process, we fuse the visual, semantic and views information into a regularization framework to learn the relevance score of every image in each user’s image set. To speed up the learning speed, we use the co-occurrence word set of the given query to estimate the semantic relevance matrix.

### 5.3 MODULES DESCRIPTION

#### LIST OF MODULES:

A module is a part of a program. Programs are composed of one or more independently developed modules that are not combined until the program is linked. A single module can contain one or several routines.

Our project modules are given below:

#### 4.2 SOFTWARE SPECIFICATIONS

- Operating System : Windows 8
- Front End : Java
- Database : MySQL

#### IMPLEMENTATION

##### 5.1 JAVA:

Java technology is both a programming language and a platform. Java is a programming language and computing platform first released by Sun Microsystems in 1995. It is the underlying technology that powers state-of-the-art programs including utilities, games, and business applications. Java runs on more than 850 million personal computers worldwide, and on billions of devices worldwide, including mobile and TV devices.

##### JAVA PROGRAMMING LANGUAGE:

The Java programming Language evolved from a language named Oak. Oak was developed in the early nineties at Sun Microsystems as a platform-independent language aimed at allowing entertainment appliances such as video game consoles and VCRs to communicate. Oak was first slated to appear in television set-top boxes designed to provide video-on-demand services.

Just as the deals with the set-top box manufacturers were falling through, the World Wide Web was coming to life. As Oak’s developers began to recognize this trend, their focus shifted to the Internet and Web Runner, an Oak-enabled browser, was born. Oak’s name was changed to Java and Web Runner became the Hot Java web browser.

The excitement of the Internet attracted software vendors such that Java development tools from many vendors quickly became available. That same excitement has provided the impetus for a multitude of software developers to discover Java and its many wonderful features.

## JAVA PROGRAM RUNNING

The programming languages, you either compile or interpret a program so that you can run it on your computer. The Java programming language is unusual in that a program is both compiled and interpreted. With the compiler, first you translate a program into an intermediate language called Java byte codes —the platform-independent codes interpreted by the interpreter on the Java platform. The interpreter parses and runs each Java byte code instruction on the computer. Compilation happens just once; interpretation occurs each time the program is executed. The following figure illustrates how this works.

You can think of Java byte codes as the machine code instructions for the Java Virtual Machine (Java VM). Every Java interpreter, whether it's a development tool or a Web browser that can run applets, is an implementation of the Java VM.

Java byte codes help make “write once, run anywhere” possible. You can compile your program into byte codes on any platform that has a Java compiler. The byte codes can then be run on any implementation of the Java VM. That means that as long as a computer has a Java VM, the same program written in the Java programming language can run on Windows 2000, a Solaris workstation, or on an iMac.

## RUNNING ON JAVA PROGRAM

- Platform Independence - Java compilers do not produce native object code for a particular platform but rather ‘byte code’ instructions for the Java Virtual Machine (JVM). Making Java code work on a particular platform is then simply a matter of writing a byte code interpreter to simulate a JVM. What this all means is that the same compiled byte code will run unmodified on any platform that supports Java.
- Object Orientation - Java is a pure object-oriented language. This means that everything in a Java program is an object and everything is descended from a root object class.
- Rich Standard Library - One of Java's most attractive features is its standard library. The Java environment includes hundreds of classes and methods in six major functional areas.
  - Language Support classes for advanced language features such as strings, arrays, threads, and exception handling.
  - Utility classes like a random number generator, date and time functions, and container classes.
  - Input/output classes to read and write data of many types to and from a variety of sources. □ Networking classes to allow inter-computer communications over a local network or the Internet.

## THE JAVA PLATFORM:

A platform is the hardware or software environment in which a program runs. We've already mentioned some of the most popular platforms like Windows 2000, Linux, Solaris, and MacOS. Most platforms can be described as a combination of the operating system and hardware.

The Java platform differs from most other platforms in that it's a software-only platform that runs on top of other hardware-based platforms.

The Java platform has two components:

- The Java Virtual Machine (Java VM)
- The Java Application Programming Interface (Java API)

You've already been introduced to the Java VM. It's the base for the Java platform and is ported onto various hardware-based platforms. The Java API is a large collection of ready-made software components that provide many useful capabilities, such as graphical user interface (GUI) widgets. The Java API is grouped into libraries of related classes and interfaces; these libraries are

known as packages. The next section, What Can Java Technology Do? Highlights what functionality some of the packages in the Java API provide. The following figure depicts a program that's running on the Java platform. As the figure shows, The Java API and the virtual machine insulate the program from the hardware.

## JAVA PLATFORM

Native code is code that after you compile it, the compiled code runs on a specific hardware platform. As a platform-independent environment, the Java platform can be a bit slower than native code. However, smart compilers, well-tuned interpreters, and just-in-time byte code compilers can bring performance close to that of native code without threatening portability.

## JAVA TECHNOLOGY

We can't promise you fame, fortune, or even a job if you learn the Java programming language. Still, it is likely to make your programs better and requires less effort than other languages. We believe that Java technology will help you do the following:

- Get started quickly: Although the Java programming language is a powerful object-oriented language, it's easy to learn, especially for programmers already familiar with C or C++.
- Write less code: Comparisons of program metrics (class counts, method counts, and so on) suggest that a program written in the Java programming language can be four times smaller than the same program in C++.
- Write better code: The Java programming language encourages good coding practices, and its garbage collection helps you avoid memory leaks. Its object orientation, its JavaBeans component architecture, and its wide-ranging, easily extendible API let you reuse other people's tested code and introduce fewer bugs.
- Develop programs more quickly: Your development time may be as much as twice as fast versus writing the same program in C++. Why? You write fewer lines of code and it is a simpler programming language than C++.
- Avoid platform dependencies with 100% Pure Java: You can keep your program portable by avoiding the use of libraries written in other languages. The 100% Pure Java™ Product Certification Program has a repository of historical process manuals, white papers, brochures, and similar materials online.
- Write once, run anywhere: Because 100% Pure Java programs are compiled into machine-independent byte codes, they run consistently on any Java platform.
- Distribute software more easily: You can upgrade applets easily from a central server.

Applets take advantage of the feature of allowing new classes to be loaded "on the fly," without recompiling the entire program.

## OPEN DATABASE CONNECTIVITY

ODBC aims to provide a common API for access to SQL1-based database management systems (DBMSs) such as MySQL, PostgreSQL, Microsoft Access and SQL Server, DB2, Oracle and SQLite. It originated on Windows in the early 1990s, but ODBC driver managers on Unix, ODBC and iODBC are nowadays available on a wide range of platforms (and a version of iODBC ships with recent versions of Mac OS X).

The connection to the particular DBMS needs an ODBC driver: these may come with the DBMS or the ODBC driver manager or be provided separately by the DBMS developers, and there are third-party developers such as Actual Technologies, Easysoft and OpenLink. (This means that for some DBMSs there are several different ODBC drivers available, and they can behave differently.) Microsoft provides drivers on Windows for non-SQL database systems such as dBase and FoxPro, and even for Lotus and Excel spreadsheets. Actual Technologies sell a driver for Mac OS X that covers (some) Excel spreadsheets and flat files.

The ODBC system files are not installed on your system by Windows 95. Rather, they are installed when you setup a separate database application, such as SQL Server Client or Visual Basic 4.0. When the ODBC icon is installed in Control Panel, it uses a file

called ODBCINST.DLL. It is also possible to administer your ODBC data sources through a stand-alone program called ODBCADM.EXE.

From a programming perspective, the beauty of ODBC is that the application can be written to use the same set of function calls to interface with any data source, regardless of the database vendor. The source code of the application doesn't change whether it talks to Oracle or SQL Server. We only mention these two as an example.

There are ODBC drivers available for several dozen popular database systems. Even Excel spreadsheets and plain text files can be turned into data sources. The operating system uses the Registry information written by ODBC Administrator to determine which low-level ODBC drivers are needed to talk to the data source (such as the interface to Oracle or SQL Server). The loading of the ODBC drivers is transparent to the ODBC application program.

The advantages of this scheme are so numerous that you are probably thinking there must be some catch. The only disadvantage of ODBC is that it isn't as efficient as talking directly to the native database interface. ODBC has had many detractors make the charge that it is too slow. Microsoft has always claimed that the critical factor in performance is the quality of the driver software that is used. In our humble opinion, this is true.

The availability of good ODBC drivers has improved a great deal recently. And anyway, the criticism about performance is somewhat analogous to those who said that compilers would never match the speed of pure assembly language. Maybe not, but the compiler (or ODBC) gives you the opportunity to write cleaner programs, which means you finish sooner. hile, computers get faster every year.

#### **JAVA DATABASE CONNECTIVITY:**

In an effort to set an independent database standard API for Java; Sun Microsystems developed Java Database Connectivity, or JDBC. JDBC offers a generic SQL database access mechanism that provides a consistent interface to a variety of RDBMSs. This consistent interface is achieved through the use of "plug-in" database connectivity modules, or drivers. If a database vendor wishes to have JDBC support, he or she must provide the driver for each platform that the database and Java run on.

#### **Some Swing Components of Java:**

##### **JScrollPane:**

Provides a scrollable view of a lightweight component. It is used to display a child component with a built-in scrolling facility. The scrolling of a child component, when its size is larger than the available view port, is performed in horizontal or vertical directions by using the scrollbars associated with the scroll pane. Scroll panes are very easy to implement because the adjustment events fired by the scrollbars are already taken care of by the scrollpane object. A Swing scroll pane is an object of type JScrollPane that extends from the class JComponent.

##### **JButton:**

Swing buttons are represented by the objects of class JButton, and each button is basically an implementation of a push-type button. Unlike AWT buttons, Swing buttons can be displayed with text labels as well as icons. We can also set different icon for different states of the buttons by using supporting methods.

##### **JFrame:**

An extended version of java.awt.Frame that adds support for the JFC/Swing components architecture. The JFrame class is slightly incompatible with Frame. Like all other JFC/Swing toplevel containers, a JFrame contains a JRootPane as its only child. The content pane provided by the root pane should, as a rule, contain all the non-menu components displayed by the JFrame.

##### **JTextField:**

The swing text field can be used to display or edit a single line of plain text. The component appears similar to the AWT text field; however, the Swing text field is a lightweight component. A text-field object is created by using the class JTextField, which is a direct

subclass of JComponent. Thus, the functionality of JTextField spreads into JComponent and JComponent. JTextField objects can fire the action and mouse events that can be captured by a registered listener.

## 5.2 MS-SQL SERVER 2000

MS SQL Server is a powerful database management system and the user can create applications that require little or no programming. It supports GUI features and an entire programming language, Visual Studio Application which can be used to develop richer and more developed applications.

There are quite a few reasons, the first being that SQL is a feature rich program that can handle any database related task you have. You can create places to store your data build tools that make it easy to read and modify your database contents, and ask questions of your data. SQL is a relational database, a database that stores information about related objects. In MS SQL that database means a collection of tables that hold data. It collectively stores all the other related objects such as queries, forms and reports that are used to implement function effectively.

The MS SQL database can act as a back end database for .NET as a front end, MS SQL supports the user with its powerful database management functions. A beginner can create his/her own database very simply by some mouse clicks. Another good reason to use SQL as backend tool is that it is a component of the overwhelmingly popular Microsoft office software suite.

### 1. SQL LEVEL API

The designers felt that their main goal was to define a SQL interface for Java. Although not the lowest database interface level possible, it is at a low enough level for higher-level tools and APIs to be created. Conversely, it is at a high enough level for application programmers to use it confidently. Attaining this goal allows for future tool vendors to “generate” JDBC code and to hide many of JDBC’s complexities from the end user.

### 2. SQL CONFORMANCE

SQL syntax varies as you move from database vendor to database vendor. In an effort to support a wide variety of vendors, JDBC will allow any query statement to be passed through it to the underlying database driver. This allows the connectivity module to handle non-standard functionality in a manner that is suitable for its users.

### 3. JDBC MUST BE IMPLEMENTAL ON TOP OF COMMON DATABASE INTERFACES

The JDBC SQL API must “sit” on top of other common SQL level APIs. This goal allows JDBC to use existing ODBC level drivers by the use of a software interface. This interface would translate JDBC calls to ODBC and vice versa.

### 4. PROVIDE A JAVA INTERFACE THAT IS CONSISTENT WITH THE REST OF THE JAVA SYSTEM

Because of Java’s acceptance in the user community thus far, the designers feel that they should not stray from the current design of the core Java system.

### 5. KEEP IT SIMPLE

This goal probably appears in all software design goal listings. JDBC is no exception. Sun felt that the design of JDBC should be very simple, allowing for only one method of completing a task per mechanism. Allowing duplicate functionality only serves to confuse the users of the API.

### 6. USE STRONG, STATIC TYPING WHEREVER POSSIBLE

Strong typing allows for more error checking to be done at compile time; also, less error appear at runtime.

### 7. KEEP THE COMMON CASES SIMPLE

The more often than not, the usual SQL calls used by the programmer are simple



SELECT's, INSERT's, DELETE's and UPDATE's, these queries should be simple to perform with JDBC. However, more complex SQL statements should also be possible.

### FUNCTIONALITY:

JDBC allows multiple implementations to exist and be used by the same application. The API provides a mechanism for dynamically loading the correct Java packages and registering them with the JDBC Driver Manager. The Driver Manager is used as a connection factory for creating JDBC connections.

JDBC connections support creating and executing statements. These may be update statements such as SQL's CREATE, INSERT, UPDATE and DELETE, or they may be query statements such as SELECT. Additionally, stored procedures may be invoked through a JDBC connection. JDBC represents statements using one of the following classes:

- Statement – the statement is sent to the database server each and every time.

- Prepared Statement – the statement is cached and then the execution path is pre determined on the database server allowing it to be executed multiple times in an efficient manner.
- Callable Statement – used for executing stored procedures on the database.

Update statements such as INSERT, UPDATE and DELETE return an update count that indicates how many rows were affected in the database. These statements do not return any other information.

Query statements return a JDBC row result set. The row result set is used to walk over the result set. Individual columns in a row are retrieved either by name or by column number. There may be any number of rows in the result set.

### MY SQL:

MySQL develops all of the documentation internally. This includes the various versionspecific reference manuals, the documentation for the GUI tools such as MySQL Administrator and MySQL Query Browser, and specialized guides such as the NDB API Internals manual. These manuals are provided in multiple formats and are accessible from the main MySQL Documentation page.

MySQL is an open source relational database management system (RDBMS). It is commonly used for database services for other open source applications, such as Drupal and Word press. It provides two methods to use MySQL for customer applications. In the first method, ITS fully manages the entire MySQL infrastructure so the customer does not have to. In the second method, the customers manage the infrastructure themselves and purchase access to the university's campus-wide MySQL Enterprise Platinum support contract.

This guide provides information on the format, structure, and builds processes that are used to develop and produce the documentation, and information on how you can build your own custom documentation using the publicly available Subversion repository. Note, though, that publishing MySQL documentation is subject to the limitations stated in our legal notice. It also contains a style guide on writing proper documentation, and a DocBook guide covering mark-up. All of the MySQL documentation is written using the DocBook XML format. Using DocBook XML enables us to easily generate the documentation in a number of different output formats, including.

MySQL is a relational database management system (RDBMS) based on SQL (Structured Query Language). First released in January, 1998, MySQL is now one component of parent company MySQL AB's product line of database servers and development tools.

Internet startups became interested in the original open source version of MySQL as an alternative to the proprietary database systems from Oracle, IBM, and Informix. MySQL is currently available under two different licensing agreements: free of charge, under the GNU General Public License (GPL) open source system or through subscription to MySQL Network for business applications.

MySQL runs on virtually all platforms, including Linux, UNIX, and Windows. It is fully multi-threaded using kernel threads, and provides application program interfaces (APIs) for many programming languages, including C, C++, Eiffel, Java, Perl, PHP, Python, and Tcl.

In addition to the standard DocBook XML source and transformations into these formats, we have built custom templates and translations to generate custom markup, and a number of tools and utilities that provide enhanced functionality.

The MySQL documentation repository is called `mysqldoc`. The documentation tools are in the `mysqldoc-toolset` repository, though. To work with MySQL documentation (rather than just viewing it), you need the documentation tools, too.

Both repositories, `mysqldoc` and `mysqldoc-toolset`, can be downloaded in `tgz` (tarred zip) format. The repositories contain a combination of the actual documentation source and the components and systems required to build the documentation, including tools for related tasks such as validating the XML or counting words.

## SYSTEM STUDY

### 6.1 FEASIBILITY STUDY

It is to determine the possibility of either improving the existing system or develop a totally new system. The phrase says, something that is feasible is something that can be accomplished.

The given content is competitive with the potential solution.

- Technical Feasibility
- Economic Feasibility
- Operational Feasibility

#### 6.1.1 TECHNICAL FEASIBILITY

Technical feasibility is concerned with such equations as whether easy technology exists to do what is planned. Those the proposed equipment have the data required est.'s far as a project is concerned it is technically

#### 6.1.2 ECONOMIC FEASIBILITY

In this user examine the cost of developing and operating the system with regard to what the organization can effort the project can be done on PC with the hard disk. A well done Feasibility study is firm to avoid common mistakes often made in the system.

- ❖ Failure to clearly specify problem and objectives.
- ❖ Estimate error.
- ❖ The hardware approaches etc.

The cost to conduct a full system investigation

1. The cost of hardware and software for the class of application being considered.
2. The development cost.
3. The cost maintenance etc.

### 6.1.3 OPERATIONAL FEASIBILITY

Will the system be used if it is developed and implemented? Will there be resistance from users that will undermine the possible application benefits? The feasibility study is carried out by a small group of people who are familiar with information systems techniques, understand the parts of the business or organization that will involve or be affected by the project, and are skilled in the system analysis and design process.

## 6.2 SYSTEM ANALYSIS

In this chapter, concepts associated with term structured system and how they are implemented in the project has been dealt with the tools used for structure system analysis are

- Data Flow Diagram
- Entity Relationship Diagram

### 6.2.1 ENVIRONMENTAL MODEL

The Environmental model defines the interfaces between the system and the environment. Building an environmental model is the first and the most important part of building complete model of user requirements. The critical aspect of environmental model is to identify the events occurring in the environment to which the system must respond. It also defines the boundary between the system and the environment.

### 6.2.2 BEHAVIOURAL MODEL

A data flow diagram is a pictorial representation of the working of the system. It is a tool that depicts the flow of data through a system and the work processing performed by that system. This takes an important role in the system analysis part to know the present level of existing system and what modification is to be done to overcome the problem occurring in the system. It is the starting point of the design phase that functionally decomposed the requirement. A data flow diagram consists of a series of bubbles joined by lines. The bubbles represent data transformation and lines represent data flow in those systems.

### 6.2.3 DATA MODEL

Entity Relationship Diagram is a model that describes the stored data layout of a system at high level of abstraction. Entity Relationship Diagram enables us to examine and highlight data structures and relationship independent of the processes or control flow. The participation of the entities in the relationship and cardinality ratio are indicated in the entity relationship diagram to provide further information. The Entity Relationship Diagram shows the relationship and types of relationship between the entities. The entity for the system. The major components are

- ❖ *Rectangle* - which represent the entity sets
- ❖ *Ellipses* - which represents attributes
- ❖ *Diamond* - which represent relationship set
- ❖ *Line* - which links the attributes to entity sets and entity sets to relationship set

### DATA FLOW DIAGRAM (DFD)

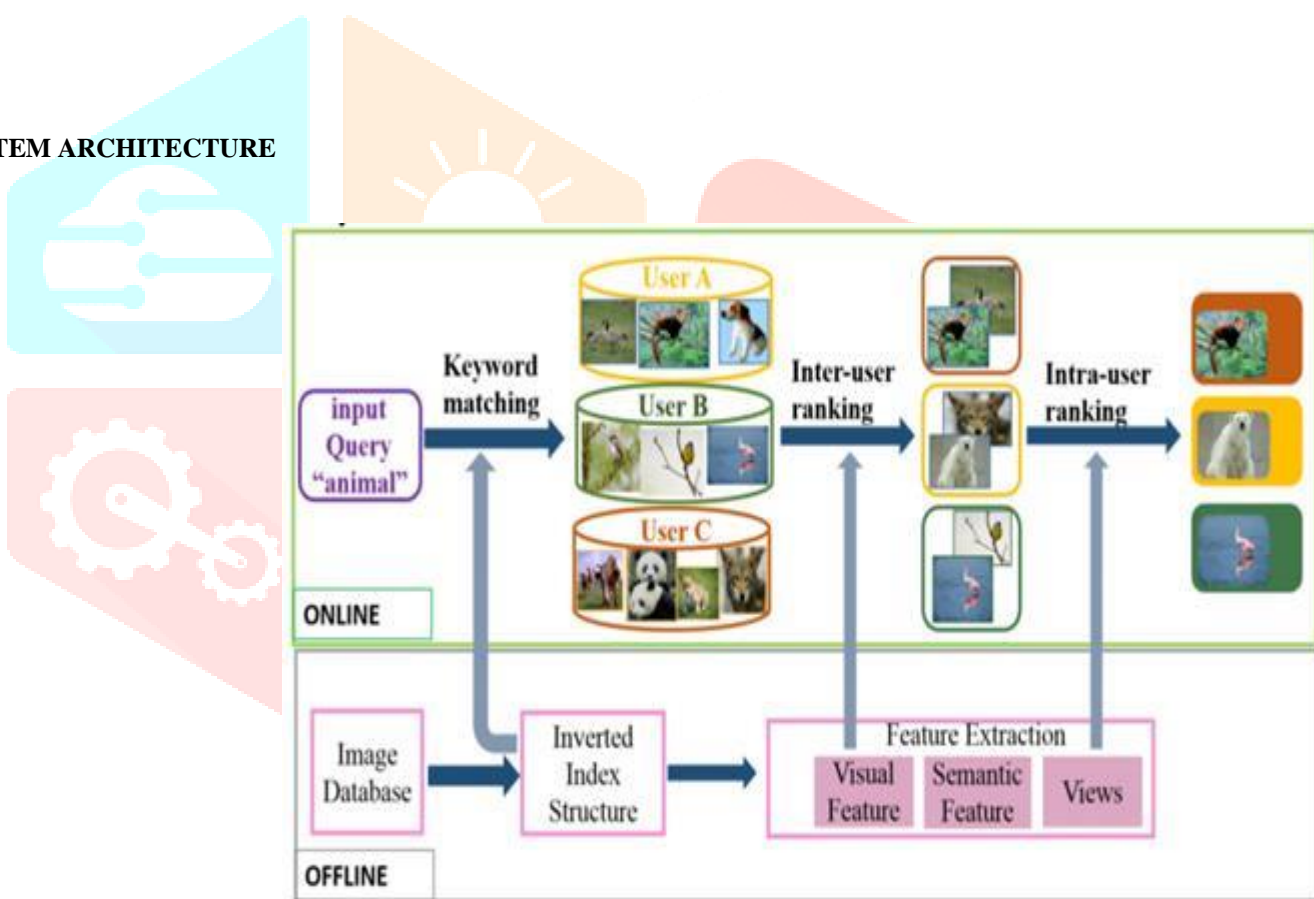
The first step is to draw a data flow diagram (DFD). The DFD was first developed by Larry Constantine as a way of expressing system requirements in graphical form.

A DFD also known as a “bubble chart” has the purpose of clarifying system requirements and identifying major transformations that will become programs in system design. So, it is the starting point of the design phase that functionally decomposes the requirements specifications down to the lowest level of detail. A DFD consists of series of bubbles join by the data flows in the system.

The purpose of data flow diagrams is to provide a semantic bridge between users and systems developers. The diagrams are:

- Graphical, eliminating thousands of words;
- Logical representations, modeling WHAT a system does, rather than physical models showing HOW it does it;
- Hierarchical, showing systems at any level of detail; and
- jargon less, allowing user understanding and reviewing.

### 6. SYSTEM ARCHITECTURE



6.3.3.1 System Architecture

### DATABASE DESIGN

A well database is essential for the good performance of the system .several tables are referenced or manipulated at various instance. The table, also know as relation; provide information pertaining to a specified entity. Normalization of table is carried out to extent possible, while the normalizing tables, care should be taken to make sure that the number of tables do not exceed the optimum level, so that table maintenance. Is convenient and effective

The process of doing database design generally consists of a number of steps which will be carried out by the database designer. Not all of these steps will be necessary in all cases. Usually, the designer must:

- Determine the data to be stored in the database
- Determine the relationships between the different data elements
- Superimpose a logical structure upon the data on the basis of these relationships.

Within the relational model the final step can generally be broken down into two further steps that of determining the grouping of information within the system, generally determining what are the basic objects about which information is being stored, and then determining the relationships between these groups of information, or objects. This step is not necessary with an Object database.

In a majority of cases, the person who is doing the design of a database is a person with expertise in the area of database design, rather than expertise in the domain from which the data to be stored is drawn e.g. financial information, biological information etc. Therefore the data to be stored in the database must be determined in cooperation with a person who does have expertise in that domain, and who is aware of what data must be stored within the system.

### **TABLE DESIGN**

Database design is a collection of interactive data store. It is an effective method of defining, storing and retrieving the information in the database. The database design is independent of any relational database management system and it is a logical model. The logical design is mapped according to RDBMS used for implementation.

The data contained in the database can be multiple application and users. It prevents the unauthorized from accessing data and ensures the privacy of data. The ERD to relational scheme mapping is done in order to make the most efficient use of table space. The relations of the scheme are converted into table and key attributes are converted into primary keys. The various tables that are used in the system are derived from the Entity Relationship Diagram. The tables maintained are customer table, purchase table, loan table, sales table, cash table, registration table, insurance table, employee pay slip table.

### **INPUT DESIGN**

Input design is the process of converting the user-oriented. Input to a computer based format. The goal of the input design is to make the data entry easier, logical and free error. Errors in the input data are controlled by the input design. The quality of the input determines the quality of the system output.

All the data entry screen are interactive in nature, so that the user can directly enter into data according to the prompted messages. The user are also can directly enter into data according to the prompted messages. The users are also provided with option of selecting an appropriate input from a list of values. This will reduce the number of error, which are otherwise likely to arise if they were to be entered by the user itself.

Input design is one of the most important phase of the system design. Input design is the process where the input received in the system are planned and designed, so as to get necessary information from the user, eliminating the information that is not required. The aim of the input design is to ensure the maximum possible levels of accuracy and also ensures that the input is accessible that understood by the user.

Input design features can ensure the reliability of the system and produce result from accurate data or they can result in the production of erroneous information.

## OUTPUT DESIGN

Output design is very important concept in the computerized system, without reliable output the user may feel the entire system is unnecessary and avoids using it. The proper output design is important in any system and facilitates effective decision-making. The output design of this system includes various reports.

Output requirements are designed during system analysis. A good starting point for the output design is the data flow diagram. Human factors reduce issues for design involved addressing internal controls to ensure readability.

An application is successful only when it can provide efficient and effective reports. Reports are actually presentable form of the data. The report generation should be useful to the management for future reference. The report are the main source of information for user's operators and management. Report generated are a permanent record of the transaction occurred. After any valid transactions; have commenced the report of the same are generation and: filed for future reference. Great care has been taken when designation the report as it plays an important role in decisionmarking.

## TESTING AND IMPLEMENTATION

### 7.1 TESTING

Testing is a series of different tests that whose primary purpose is to fully exercise the computer based system. Although each test has a different purpose, all work should verify that all system element have been properly integrated and performed allocated function. Testing is the process of checking whether the developed system works according to the actual requirement and objectives of the system.

The philosophy behind testing is to find the errors. A good test is one that has a high probability of finding an undiscovered error. A successful test is one that uncovers the undiscovered error. Test cases are devised with this purpose in mind. A test case is a set of data that the system will process as an input. However the data are created with the intent of determining whether the system will process them correctly without any errors to produce the required output.

### SYSTEM TESTING

It is the stage of implementation, which ensures that system works accurately and effectively before the live operation Commences. It is a confirmation that all are correct and opportunity to show the users that the system must be tested with text data and show that the system will operate successfully and produce expected results under expected conditions.

Before implementation, the proposed system must be tested with raw data to ensure that the modules of the system work correctly and satisfactorily. The system must be tested with valid data to achieve its objective.

The purpose of system testing is to identify and correct errors in the candidate system. As important as this phase is, it is one that is frequently compromised. Typically, the project the schedule or the user is eager to go directly to conversion. Actually, testing is done to achieve the system goal. Testing is vital to the parts of the system are correct; the goal will be successfully achieved. Inadequate testing or non-testing leads to errors that may not appear until months later Appearance of the problem.

## Types of Testing:

- Unit testing
- Integration testing
- Validation testing
- Output testing
- User acceptance testing

### 7.1.1 Unit Testing

All modules were tested and individually as soon as they were completed and were checked for their correct functionality. Unit testing involves only those characteristics that are vital to the performance of the unit under test. This encourages developers to modify the source code without immediate concerns about how such changes might affect the functioning of other units or the program as a whole. Once all of the units in a program have been found to be working in the most efficient and error-free manner possible, larger components of the program can be evaluated by means of integration testing.

### 7.1.2 Integration Testing

The entire project was split into small program; each of this single programs gives a frame as an output. These programs were tested individually; at last all these programs were combined together by creating another program where all these constructors were used. It give a lot of problem by not functioning is an integrated manner.

The user interface testing is important since the user has to declare that the arrangements made in frames are convenient and it is satisfied. when the frames were given for the test, the end user gave suggestion. Based on their suggestions the frames were modified and put into practice.

### 7.1.3 Validation Testing

At the culmination of the black box testing software is completely assembled as a package. Interfacing errors have been uncovered and corrected and a final series of test i.e., Validation succeeds when the software function in a manner that can be reasonably accepted by the customer.

### 7.1.4 Output Testing

After performing the validation testing the next step is output testing of the proposed system. Since the system cannot be useful if it does not produce the required output. Asking the user about the format in which the system is required tests the output displayed or generated by the system under consideration. Here the output format is considered in two ways. one is on screen and another one is printed format. The output format on the screen is found to be corrected as the format was designed in the system phase according to the user needs. And for the hardcopy the output comes according to the specifications requested by the user.

### 7.1.5 User Acceptance System

An acceptance test as the objective of selling the user on validity and reliability of the system. It verifies that the procedures operate to system specification and mat the integrity of vital is maintained.

### 7.1.6 Performance Testing

This project is a application based project, and the modules are interdependent with the other modules, so the testing cannot be done module by module. So the unit testing is not possible in the case of this driver. So this system is checked only with their performance to check their quality.

### 7.2 IMPLEMENTATION

System implementation is stage in the project where the theoretical design is turned into the working system. The most crucial stage is giving the users confidence that the new system will work effectively and efficiently.

The performance of reliability of the system is tested and it gained acceptance. The system was implemented successfully. Implementation is a process that means converting a new system in to operation.

Proper implementation is essential to provide a reliable system to meet organization requirements. During the implementation stage a live demon was undertaken and made in front of end-users. The various features provided in the system were discussed during implementation **CHAPTER 8**

### CONCLUSION

The propose a social re-ranking method for tag-based image retrieval. In this social reranking method, inter-user re-ranking and intra-user re-ranking are carried out to obtain the retrieved results. In order to enhance the diversity performance, user information is firstly introduced into our proposed approach and obtains satisfactory results. Besides, views of social image is also firstly fused into a traditional regularization framework to enhance the relevance performance of retrieved results. Discussions and experiments have demonstrated that our proposed method is effective and time-saving. However, in the inter-user ranking process only user's contribution is considered and the similarity among users is ignored. In addition to this, many information in Flickr dataset are still ignored, such as title information, time stamp and so on.

### REFERENCES

- [1] Xueming Qian, Yisi Zhao, Junwei Han: Image Location Estimation by Salient Region Matching. *IEEE Transactions on Image Processing* 24(11): 4348-4358 (2015)
- [2] Y. Yang, Y. Gao, H. Zhang, J. Shao. Image Tagging with Social Assistance. In *Proceedings of International Conference on Multimedia Retrieval* (p. 81), ACM, 2014.
- [3] A. Ksibi, AB. Ammar, CB. Amar. Adaptive diversification for tag-based social image retrieval. *International Journal of Multimedia Information Retrieval*, 2014, 3.1: 29-39.
- [4] X. Qian, H. Wang, G. Liu, X. Hou, "HWVP: Hierarchical Wavelet Packet Texture Descriptors and Their Applications in Scene Categorization and Semantic Concept Retrieval", *Multimedia Tools and Applications*, May 2012.
- [5] Y. Gao, M. Wang, H. Luan, J. Shen, S. Yan, and D. Tao. Tag-based social image search with visual-text joint hypergraph learning. *Proceedings of the ACM International Conference on Multimedia information retrieval*, 2011:1517-1520.
- [6] K. Yang, M. Wang, X. Hua, and H. Zhang. Social Image Search with Diverse Relevance Ranking. *Proceedings of the IEEE International Conference on Magnetism and Magnetic Materials*, 2010:174-184.
- [7] M. Wang, K. Yang, X. Hua, and H. Zhang. Towards relevant and diverse search of social images. *IEEE Transactions on Multimedia*, 12(8):829-842, 2010.



- [8] D. Liu, X. Hua, M. Wang, and H. Zhang. Boost Search Relevance For Tag-Based Social Image Retrieval. Proceedings of the IEEE International Conference on Multimedia and Expo, 2009:1636-1639.
- [9] D. Liu, X. Hua, L. Yang, M. Wang, and H. Zhang. Tag ranking. Proceedings of the IEEE International Conference on World Wide Web, 2009: 351-360.
- [10] R. Leuken, L. Garcia, X. Olivares, and R. Zwol. Visual diversification of image search results. In Proc. WWW Conf., 2009, pp.341–350.
- [11] X. Qian, G. Liu, D. Guo. Object categorization using hierarchical wavelet packet texture descriptors. in Proc. ISM 2009, pp.44-51.
- [12] X. Li, C. Snoek, and M. Worring. Learning tag relevance by neighbor voting for social image retrieval. Proceedings of the ACM International Conference on Multimedia information retrieval, 2008: 180-187.,
- [13] R. Cilibrasi and P. Vitanyi. The Google Similarity Distance. IEEE Transactions on Knowledge and Data Engineering, 19(3):1065-1076, 2007.
- [14] K. Song, Y. Tian, T. Huang, and W. Gao. Diversifying the image retrieval results. In Proc. ACM Multimedia Conf., 2006, pp. 707–710.
- [15] D. Cai, X. He, Z. Li, W. Ma, and J. Wen. Hierarchical clustering of WWW image search results using visual, textual and link information. In Proc. ACM Multimedia Conf., 2004, pp.

