MSc Computer Science
Project Report 2014

Customer Feedback Retrieval (Ranking and Highlighting) for Management Decision-Making

Name: Arshad khan

Student No: 120658365

Supervisor: Dr Thomas Roelleke

Date: 26 August 2014

Disclaimer

This report is submitted as part of requirement for the degree of MSc Computer Science at the Queen Mary University of London. This report has been prepared in good faith on the basis of information available. Readers are responsible for assessing the relevance and accuracy of the content of this report. This report is totally my own effort except where indicated in the report text. The report may be freely copied and distributed provided the source is acknowledged.

Signature:

Date:

Acknowledgement

I would like to express my gratitude to people who helped me producing this project and helped me making this dissertation report. I would like to express my very great appreciation to Dr.Thomson Roelleke for his valuable and constructive suggestions during the planning and development of this project report. His willingness to give his time so generously has been very much appreciated. I would also like to extend my thanks to my previous colleagues and for their help in offering me the resources in running the program. I would finally like to express my gratitude to all my friends for the their valuable feedback and suggestion during the dissertation.

Abstract

The idea behind the project is knowledge Representation by Retrieving, Ranking and highlighting the feedbacks of customer's suggestion. Rail travel is very important in our life because we use rail services every day for our travel purposes and we face different experiences in our every day travel, so every customer give some feedback about the rail systems, feedback are very help full for the top management for decision making for further planning and development. To improve company policies and improvement of the rail system for effective services it is very important to understand the user feedback regarding services provided at stations and in train. Due to uncertainties of knowledge in rail system generate a big impact on management decision-making process. In this project I will implement high level knowledge representation and Information Retrieval to provide ranking highlights consist of good suggestion and important observations using calculating probabilistic knowledge for user feedback analysis. To facilitate user feedback analysis for decision making for management a knowledge base system and Bayesian networks has developed. Bayesian networks tell us the process of calculating the probabilistic regarding user feedback analysis [14].

In this project I am using rail customer feed backs as document.

Table of Contents

Table of Tables

Table of Figures

Chapter 1: Introduction

1.1 Project Overview

Retrieving unstructured information from the document or from the online web sites have gained significant popularity and now among the most popular sites on the web like different search engines (Google, Yahoo, Bing, AltaVista etc), nearly every website have some sort of search facility to search particular data in website document. This make the concept of information retrieval more important in terms of user need to search particular data in large collection of documents where some time million of document are available and searching and looking particular data in each document manually is time consuming job.
Meaning of term "information retrieval" is very broad. Just getting credit card out from your pocket and typing credit card no is form of information retrieval. But we describe information retrieval as "Information retrieval is (IR) is finding material (documents) of an unstructured nature (text) that satisfies an information need from within large collection (store in computer).
People and industries engage in information retrieval are reference librarians, paralegals and similar professions.

1.2 Summary of Project

Rail is very important transport in the life of commuter traveling from home to work and vise versa and for going to shopping center or may be any party at some distance this all some time need to use rail travel from one end to another and using facilities available in rail and at stations in form of lifts, escalators, TVM (ticket vending machine), public information display (PID), Passenger Valuator's oyster touching machines and information available in stations regarding travel time of trains.

Commuter leaves feedback relating to services provided by the company operating rail network these feedback are very important to top management to plane for the future planning. For example customer provide feedback related to hazard on the stations so it is very good for management to find out easily what hazard are people facing related to health and safety. The management can find these all feedback using information retrieval system. But finding particular information in case of where feedbacks are in megabytes it is

difficult for management decision to create decision making skill on the basis of just retrieving results so there needed for customer feedback to interpret the data and create valuable knowledge that help finding correct customer feedback for this purpose we gain advantages of application of knowledge base system (KBS). This knowledge base system provides knowledge for solving complex user feedbacks.

1.3 Project Aims, Motivation and Objective

The aims, motivation and objectives of this customer feedback retrieval (ranking and highlighting) project are listed below.

- The main goal of this Retrieval and Ranking is to help top management to find out best suggestion given by rail user.
- To find out suggestion which make it possible to find out what main hazards are on rail service for commuter.
- To find out suggestion which are helpful to increase facilities for rail user in train and at stations.
- To find out suggestion and comments which bring rail service closer to the standard of commuter requirement.
- To find out suggestion of customers which Connect commuter to and from destination with good service.
- To find out suggestion which provide describes services at station to help customers.
- To find out suggestions those about best ticketing and oyster reader facilities at stations.
- To find information those minimize the risk of hazards at the stations and in the train during jumping in and out of the train during train service.

For all these above knowledge's derivation we need expertise knowledge in formal system where knowledge can be extracted and manipulated efficiently. Knowledge can be obtain by analysis of information in intelligent system so we need correct information for taking a right and accurate decision this knowledge is necessary to support reasoning. Process of reasoning based on rules, these rules fulfills the condition by using facts. [14]

In the case of decision making logic is on of the most popular way due to mathematical reason. So in knowledge representation classical logic used is the

first order logics. There are different types of logic used in process of reasoning like model logics, probabilistic methods and fuzzy logic that could provide some a power to reason regarding possibilities and further accurately to reason uncertain situations. In decision-making process uncertainty is main issue that decision maker face.

To overcome these uncertainties probability theory developed as solution. Bayesian network is one of the probability theory that deals with uncertainty in knowledge representation and reasoning [14].

Chapter 2: Literature Review

Decision making expert system is the branch of computer science which known as artificial intelligence (AI). Artificial intelligence basically concerned with knowledge representation, solving of problems, studying, robotics, and making of those computers, which speak and understand human like [15].
In others words we can define an expert system is a type of computer program, which reasons by using knowledge to solve the very complex problems. So these system provide the opportunity of saving and reviving human skills in very relax and acceptable way compare to traditional software using only declarative programming [15].

2.1 Background

Information retrieval is a direction that deals to show the storage, organization, and access to information items. The basic goal of information retrieval is to get information that might be useful or close relevant to the need of the any user, for example library card cabinets are a "conventional" retrieval system, and, in some ways, even searching for a visiting card in pocket to discover a colleague's details might be considered as an information retrieval task.
But knowledge base system has been developed in recent years. The purpose of knowledge base expert system for decision-making developed for much kind of applications, which involve identification, forecast, discussion, and information retrieval.
CALEX is one of the expert system, which is developed recently. This system used to find out peach and nectarine disorders. In our information retrieval of customer feedback for decision-making is used to retrieval information for management decision making at top level to decide different services and do future planning on the basis of customer feedback.

2.2 Detailed Discussion

Information retrieval could be defined as the regulation that find the relevant documents as against to easy matches to lexical patterns in a query. This highlighted a basic fundamental feature of information retrieval, that the relevance of results is calculated relative to the information requirement but not based on the query. If we describe this by taking into account the information requirement of assess out whether eating biscuit is advantageous in diminishing blood pressure. We can convey this by the search engine query as "biscuit effect blood pressure" we assess a resulting documents as relevant if it mark the information requirement, but not because of it contains similar words in the query.
This observes to be a very good relevance sign by most of information retrieval models.
It may be well known that relevance is a idea with very interesting qualities, it is subjective to two different users may have the similar information requirement but give totally different judgments regarding the similar retrieved document. Another view is its active nature; any documents retrieved and presented to the user at any given time might affect relevance judgments on the documents that will be showed after. Further according to current ranking, a user can show totally different perception regarding the similar document (on similar query) at the end relevance is many sides, it is determined not just by the content of a retrieved result but also by aspects such as the authoritativeness, probability, specificity, exhaustiveness and comprehensibility of its source. Relevance is not known to the any system before the user common sense.
We can describe main job of the information retrieval system is to "calculate" a set of documents "D" those relevant with respect to a user query "Q".
Computing a relevance R (q_k, d_j) for every document as $d_j \in D$. "R" depends on the information retrieval model. There are many retrieval models available in the market.

Search Engine is the most vital and popular application of information retrieval (IR), but IR techniques are also fundamental to the number of other tasks.

Filtering information using filtering systems remove redundant or not information from a stream using methods before displaying theses data to human eye. A model application of data filtering is the malware filters, which distinguish between useful and injurious emails depend on the native content of the emails and on the users behavior when processing them.

The Document summarization is another information retrieval (IR) application that used to create a shorter shape of an entire text in order to reduce the information load. Summarization in general is extractive for example it run by selecting the most important and relevant sentences in a document and collecting them to form a shorter version of the document.
Document clustering and categorization are also very important applications of information retrieval (IR). Clustering consists of grouping documents together, which based on their proximity in an unsupervised manner.
But categorization begins from already classification of different classes and assigns each document to most relevant class. Classic applications of text categorization are the recognition of news article categories or language, while clustering is often applied to group together dynamically made search results by using their topical similarity.

Question and answering (QA) systems deals with the selection of document, which is relevant in portions, retrieving information as answers to questions never done before, as the main characteristic of Question and answering (QA) systems is the use of the system called fine-grained relevance models, this system provide answer as relevant sentences, phrases, or even words it depend on what type of question asked.

In Information retrieval of customer feedback in rail system the main concern is with the main hazards and future planning. How top management can easily filter information using customer feedbacks. For example if top management want to spend 100 million of stations how they decided where to spend to get less complain about hazards and facilities provided or how to get best customer feedback regarding ticket vending machine.
So this system based on criteria in which no of documents are queried using different queries and information has been extracted and displayed.

Chapter 3: Introduction to Information Retrieval

This chapter introduces the concept of Information Retrieval in the filed of computer science. Information retrieval is finding the material from documents of unstructured type normally in text format that satisfies an information requirement from with in very large collection normally stored in computers. Taking wallet out from the pocket and type card number is one of the information retrieval from or type [3]. Information retrieval was the activity in which only few people were involved for example reference librarians, paralegals related people, and very similar related people are involved in information retrieval. Now computer science is developed compare to early stage and millions of people are involved in using web search means information retrieval when they use search engines or search their email. Compare to previous time information retrieval is very fast now a day by day it become the dominant form of information access. Information retrieval in computer science overtaking the traditional databases style of retrieving data in which you need to put some particular unique record to fetch all the related record for example if some one need to look for the 'Order Record' he/she must put Order_Rec _Id to retrieve that order record from database.

3.1 Concept of Information Retrieval

Information retrieval is the one of the process or activity in which we obtain information resources relevant to an information need from large collection of information resources.

Information retrieval is the process of giving answers to client on his information need. It is related to the collection, depiction, storage, organization, acquiring, operation and showing of required information needed to satisfying, information retrieval generally probe automated, computer-based, solutions. Information retrieval is not a new field, but it is a very main and demanding one, and looking suitable answer to the problems it shows is becoming more critical. Information retrieval handle with accessing and searching in unstructured information for example in text data, audio data and videos it might be from one big and large file or unroll over unconnected and various sources, in unchanged storage tools as well as on streaming data. It is part of computer science and information science and uses the techniques from statistics, database management, mathematics,

machine learning, or computational.
Modern computer science and invention of World Wide Web (1989) marked and completely permanent change to the ideal of storage, access and searching of relevant document in collections, making it possible to available to the users and indexing them for precise and full coverage retrieval [5].

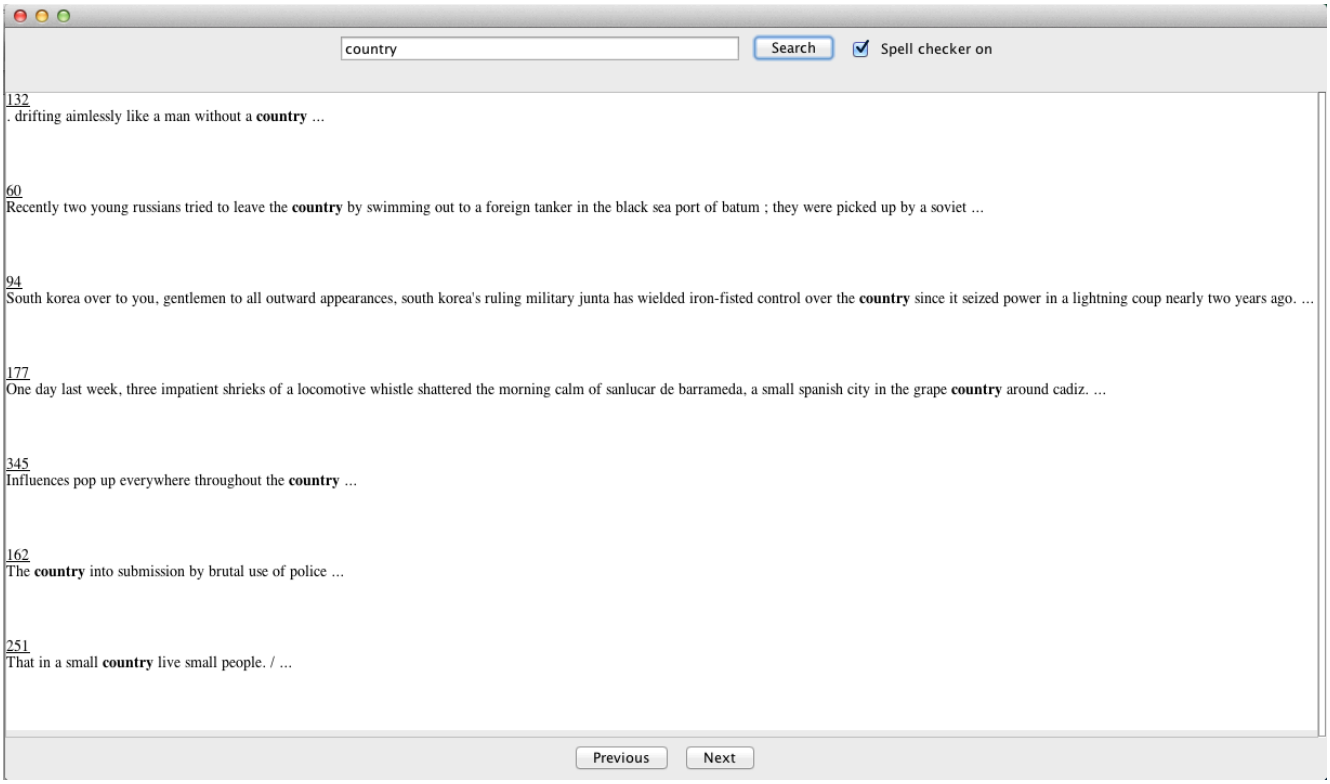

Figure 1. Search Engine Searching “Country” key word in documents

3.2 Information Retrieval Relevance

Information retrieval is defined as the field finding relevance in relevant documents as compare to simple matching of lexical patterns in a query. This mark a very basic and fundamental aspect of information retrieval and that is the relevance of result is assessed relative to the information requirement not the user query. One of the examples of this is to weather eating a mango is beneficial in reducing blood pressure. We can query in search engine as “mango effect blood pressure” so we can evaluate a document as relevant if resulting document addresses the need of information required not just it contains the all the words contain in query. Many information retrieval models have available in each of which measure relevance depending upon the algorithm implemented in that model [5].

3.3 Information Retrieval Models

Information retrieval models are developed to retrieve information. There are different types of information retrieval models some of major retrieval models are: the Boolean model, the Statistical model, in which vector space and probabilistic retrieval model, and the Linguistic and Knowledge-based models. The first model is called “exact match” model and the rest of them called “best match” models. Models are define in details in chapter 4.

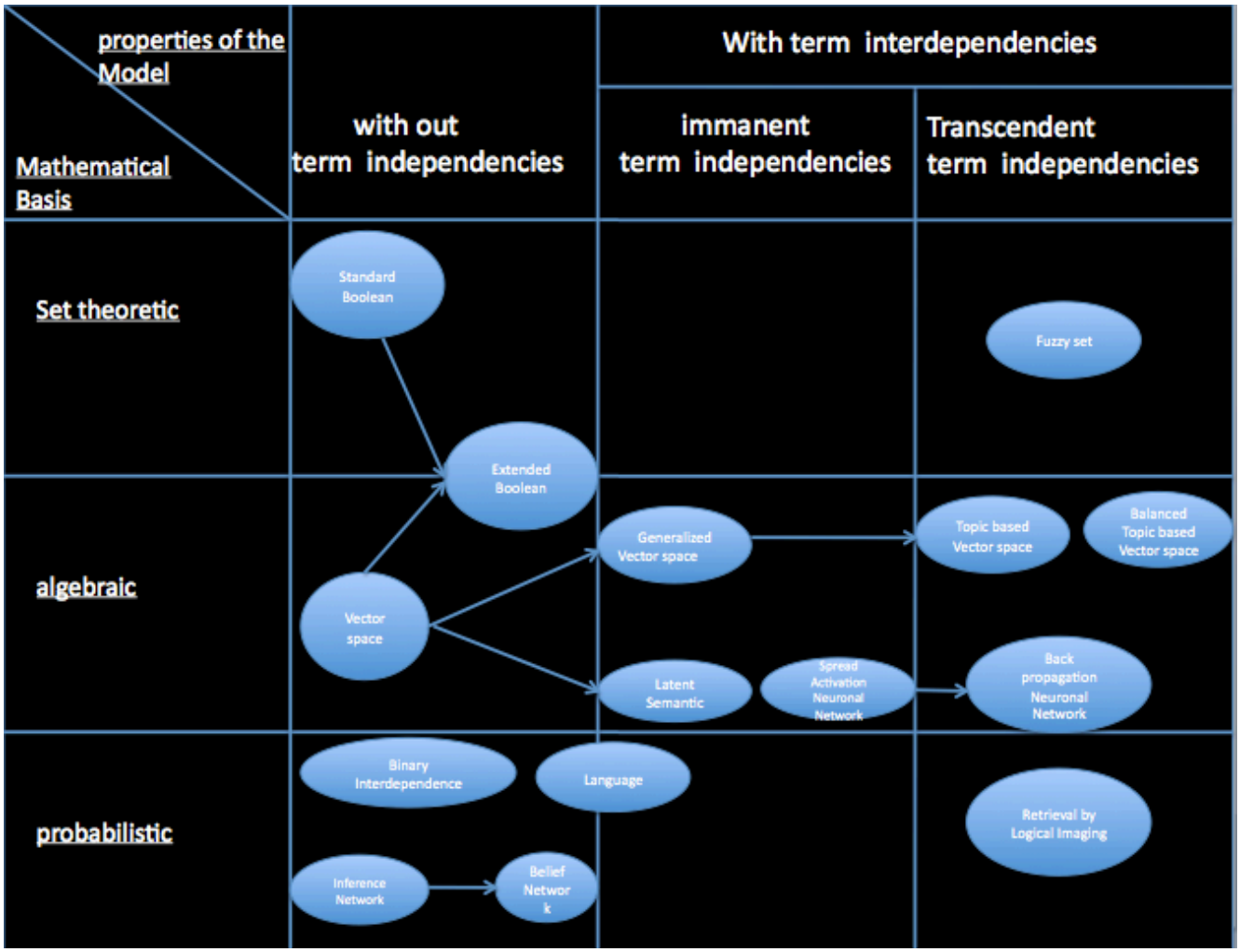

Figure 2. Information Retrieval Models [9]

3.4 Data Collection

The information retrieval as already defined is a task of finding documents those have characterized as an unstructured identity normally as text that's satisfied an information requirement need from big collection as stored on computers. As Data collection in this project is consist of around 100 documents in form of text format means data is in text format not in any other shape for example audio, video or picture Nature of Data collection is feedback from different customer traveling on rail service every day facing different kind of problems on rail network including all services like using ticket vending machines, touching oyster card machine, reading information board, using lifts services at the stations and using escalators. Some of example feedbacks are given below.

Customer Feedback Example in Text format.

"Stations have no Signs for interchanging between DLR and Tube services" 17 march 2014.

As I traveled on DLR and overall DLR experience was very good, but when I get off the Canary Wharf station I did not find any sign. We spotted one arrow that directing us from Tube to the DLR, on the way toward DLR station we met another tourist who was trying to find the canary wharf DLR station. We together ask from some workmen for directions for swapping our oyster card at DLR stations we did not find the machine to swap in our oyster card. We had to ask more people where these located in stations. Once we finally got all these, all went sorted then and all went smoothly as we need. I request please put more oyster reader and direction signs [6]

"Escalator required at Shadwell DLR stations " 20 Nov 2009

I was traveling between London Excel and Shadwell station on DLR services going to Tower Gateway station I need to jump off at shadwell station when I got off at Shadewelll station I seen a lot of people standing and waiting to go outside the station using stairs I seen one lift and that was full of people going down I realize here should be escalator to go in and out of the station because biggest hazard of this station is slipping of any person on stairs that cause major incident.

"Ticket machine not working at Bow church station" 9 April 2010

I was in hurry to go from bow church to Stratford station when I reach at bow church station I tried to buy a single ticket from machine but machine was out of order I tried another ticket machine but that was not able to process my card payment, I think this station was completely ignored by DLR operatives because we can not buy ticket when we needed in urgent.

3.5 Characterization

Information Retrieval model (IRM) is defined as a

$$IRM = \{D, Q, R(q_k, d_j), F\}$$

Where

- D is a logical set of representations of all documents in the collection of customer feedback. (Referred as d_j).
- Q is the set of representation of the user information requirement; it is called queries
  (referred as $q_{k)}$.
- R (q_k, d_j) ranking function, which relates a real number with a document representation
  d_j, which denote its relevance to a query q_k.
- F is a strategy (or framework) to model the representation of queries, document, and relationship between them.

The ranking function R (q_k, d_j) which actually declare a relevance order on the documents with contras to q_k and is a main and important element of information retrieval (IR) model [5].

3.6 Tasks in Information Retrieval

Information Retrieval (IR) most important and wider use application is search engine but techniques is very important and fundamental for number of tasks.

3.6.1 Information retrieval filtering system: is to clean repetitive and unwanted information from an information data using semi automatic method before displaying them to human eye.

3.6.2 Summarization of Documents: is one of main information retrieval (IR) application that consists in producing a very shorter version of text, this reduce the information overload this is extractive approach in which selected and most relevant sentences from a document are selected to produce the version of document.

3.6.3 Document Clustering and Categorization: is very important type of information retrieval application, the main purpose of clustering is to group document together that is based on their proximity in unsupervised fashion. But categorization is a predefined taxonomy of classes and each document is assign to relevant class. Some of the most important type of categorization is the news article category or language category.

3.6.4 Question answering: in this kind of information retrieval system deal in a manner in which it select the most relevant document portion to answer the user query. The main feature of Question answering system is fine-grained relevance models. These models answer in the form of relevant phrases, sentences and some time in even words depend on question.

3.6.5 Recommending Systems: this system is kind of information filtering in which interesting information items for example songs, movies, or books display to users based on users profiles and users close neighbor's likeness, neighbor's are selected on the basis of their proximity, social relevance, and same interest.

3.7 Information Retrieval Evaluation

To evaluate information retrieval system it is to notice that relevance is the main point to express the quality, spotlight the main fact that it direct to user requirement. It means it asses the document is relevant to user requirement means relevant or irrelevant to user query. So calculating the properties of information retrieval system, this section explain the evaluation of IR system.

3.7.1 Characteristics in Information System Evaluation

In information retrieval system there are many aspects that count toward the properties that describe the characteristics in information system evaluation.
If we see processing speed and effectiveness of document is one of convenient evaluation standard, for example by taking a component total retrieved documents in one hour and average size of these document.
Information retrieval system search speed is also very important factor because size of document collection. But main important is if we produce very fast result but result is not use full for user is like a use less information for the user need and it does not fully satisfy the user.
This means the main aspect of information retrieval system is to find the relevance, so information retrieval system can be assessable on the basis of relevance of document retrieved by user query. If document are not relevant to the user need this does not

Satisfied the user requirement and IR system is considered to be not fully functional information retrieval system. For finding the relevance of user requirement in Document collection we should have these available.

- Documents collection D.
- Set of Queries Q.

Term t = (d, q, r) where query q ∈ Q and document d ∈ D and r refers to the relevance of document d to query q [5].

3.7.2 Precision, Recall

Information retrieval system when results in disorderly it can be evaluated as precision and recall. Precision (P) and recall (R) are two basic measures used to evaluate search master plan. As if we apply search, after finishing search we first look for relevant documents or material that is important and relevant, precision and recall is the measure that use to calculate how much retrieve material is important and relevant to user need.

- **Precision:** precision can be explained, as it is ratio of relevant retrieved records with the total irrelevant and relevant records retrieved. This shows **soundness**.

$$\text{Precision (P)} = \frac{|\{\text{relevant records}\} \cap \{\text{retrieved records}\}|}{|\{\text{retrieved records}\}|}$$

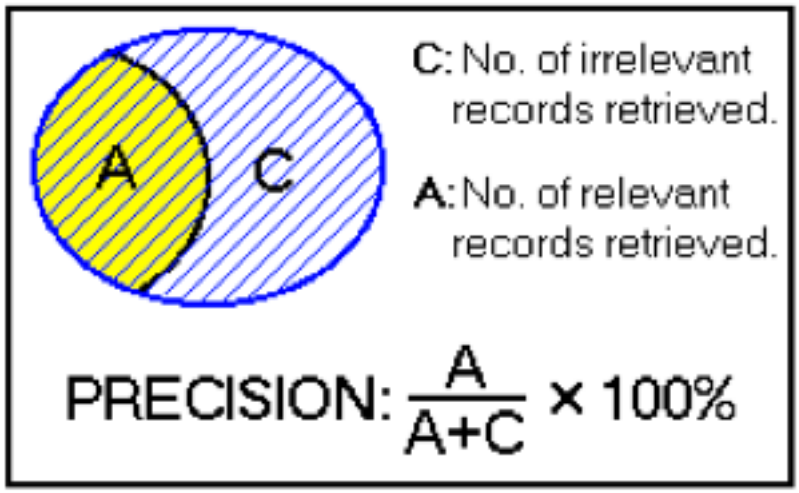

Figure 3. Precision [12]

- **Recall:** recall is totally relevant documents, it is ratio between relevant records retrieved and total relevant records in. this shows **completeness**.

$$\text{Recall (R)} = \frac{|\{\text{relevant records}\} \cap \{\text{retrieved records}\}|}{|\{\text{relevant records}\}|}$$

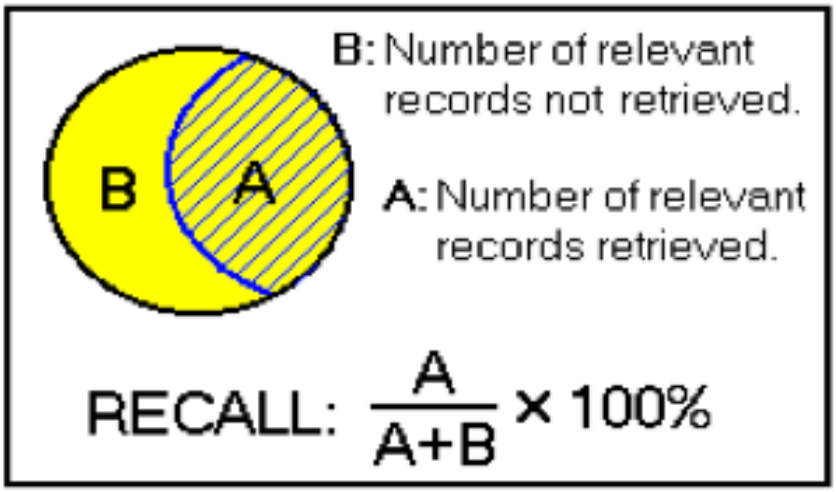

Figure 4. Recall [12]

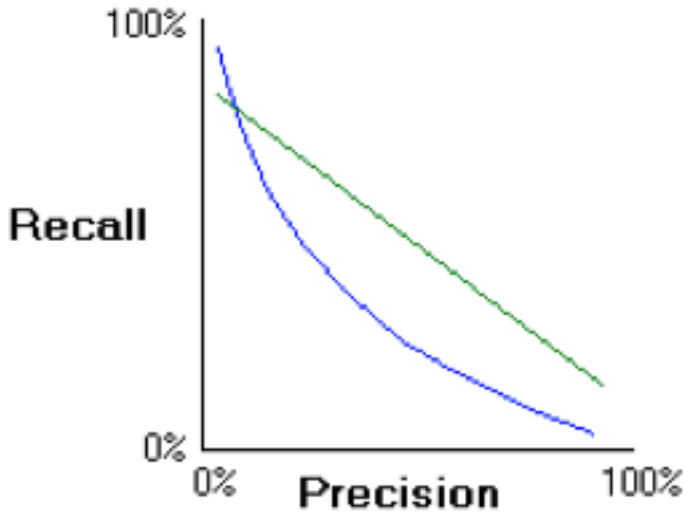

Figure 5. Precision and Recall Plot [12]

In this graph above two lines shows the performance between two different information retrieval systems in which line slop is varying.

Calculation:

More precisely if we need to understand what is precision and recall we need to solve problem as given below:

Problem:

For example

- 100 documents on particular topic.
- Query retrieved 50 documents.
- Only 35 are relevant record from 50.

Score calculation of precision and recall search is as.
X = relevant record retrieved.
Y= relevant record not retrieved.
Z = irrelevant record retrieved.

X = 35, Y = (100 – 35) = 65, Z = (50 – 35) = 15

Precision = (35/(35 + 15)) 100% => 35/50*100% = 70%

Recall = (35/(35+65))* 100% => 35/100*100% = 35%

	Google	AltaVista	HotBot	Scirus	Bioweb
Precision	0.29	0.27	0.28	0.57	0.14
Recall	0.20	0.18	0.29	0.32	0.05

Table 1. Mean Precision and Recall of Different Search Engines in 2004. [13]

3.7.3 Ranked Retrieval in Information Retrieval System

As we have seen in above section in which precision and recall is explained in detailed, as we know precision and recall is computed for unordered list of documents. What happened when evaluate the ranked retrieval result, it need to be expand these features for rank retrieval because in rank retrieval documents are only in rank form. But it does not like to be all documents only set of documents is retrieved to present. In ranked retrieval documents are accurately select to increase precision

Chapter 4: Information Retrieval Models

Information retrieval technology is very important when some one dealing with information on system or Internet. It is not possible to find information on World Wide Web with out any search Engine. With out spam filtering we cannot survive to manage our emails. For all these issues there are formal models and these models use to developed the information retrieval tools. So we are trying to understand the Information retrieval we should understand first what is retrieval model. In this chapter important retrieval models are describes.

4.1 Similarity and Matching Strategies

The fundamental and basic matter is the way for discovering the level of relevance of the user query with documents description that is known as matching process. In this process it produce a ranked list of many documents and documents those are relevant come on top of the list of ranked documents, this decrease the time to recognize required relevant documents. In other words a model of information retrieval forecasts and describe what a user find relevant, by giving his/her query. Information retrieval models work as blueprint for implementation a real information retrieval system. [4].

4.2 Boolean Model

One of the basic retrieval models in information retrieval models it is simple model based completely on the set theory and Boolean Algebra (Web Information Retrieval). It is first model in information retrieval (IR) and more criticized model. In this model queries are in the form of as Boolean expressions on index terms using operator (OR, NOT, AND) as an example “social AND economics “, “social OR political”.
Logical product is called AND as well as logical sum called OR and last one NOT is logical difference. If we see each logic for example AND like “social AND economics” bring all those documents set that is equal or less then the document sets of any term. So query will produce social and economic indexed set of documents including social and economic. If we use OR logic it combine the terms and document set will be bigger than or equal to the any one term. So using OR logic social or term economic produce the document set of either term, it explain it as union of both. As shown in figures below. Boolean model quickly define why a document is retrieved, it also clear which logical operator will bring which result either it is bigger or smaller set.
Boolean model have disadvantages that is it does not give any ranking of retrieved result. It only retrieves result or not that produces decisions, which go to annoyance .for example query leader AND worker AND union can not generate a result indexed which contain marriage, food and sweet. [4]

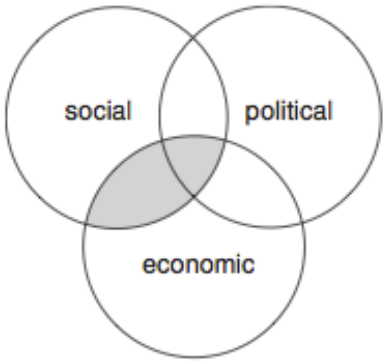

Figure 6. Economic AND Social

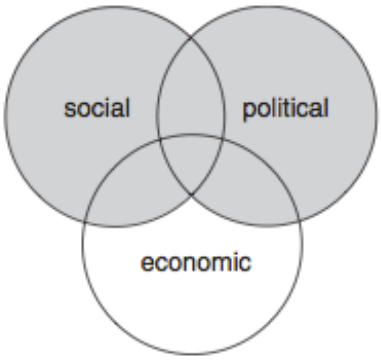

Figure 7. Political or social

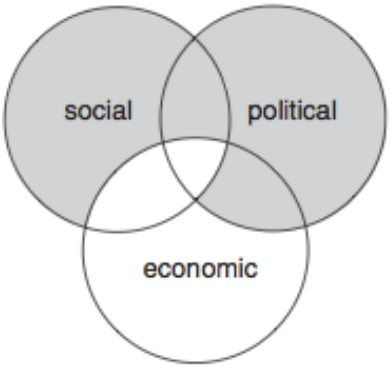

Figure 8. Political or Social AND Not (Economics or Social)

4.2.1 Boolean Retrieval

In Boolean model we can write query 'Q' in disjunctive normal form (DNF), for example

$$\mathbf{Q = term_a \wedge (term_b \vee \neg term_c)}$$

Or we can write as:
Any documents is forecast as relevant to a user query if it is satisfied the query 'Q' assertion.

$$\mathbf{((book \vee title) \wedge space \wedge \neg theory)}$$

In above Boolean model query expression each term determine a set of documents consisting the term.

Table 2. Boolean Model Query Result.

Space book
Book title
Book title: Modern space theory
Space compression

Example:

"Drinking cold milk"

(drink AND cold AND milk)

Table 3

Milk cold
Milk hot
Milk in a glass
Five days old
Me like it hot
He like it cold
Two days old
Some like it cold

4.3 Vector Space Model

In vector space model or we can call, as term vector model is a kind of algebraic model for representation of documents of type text. As name suggests both terms and queries are represented in this model as vectors in a space and each element represent a term.

In vector space model document and queries are shown as vectors as shown below

$$D_j = (W_{1,j}, W_{2,j}, W_{3,j}, W_{4,j}, \ldots\ldots\ldots\ldots\ldots\ldots , W_{t,j})$$

$$Q_j = (W_{1,q}, W_{2,q}, W_{3,q}, W_{4,q}, \ldots\ldots\ldots\ldots\ldots\ldots , W_{n,q})$$

In above every dimension represent a separate term, if document contains a term then term value in vector is greater then zero. Term weighting is one of the method to compute term values and it is also called tf * idf weighting scheme, we will see tf*idf in details in next topic.

In vector space model relevant documents can be computed by using similarity of documents in which model compare the deviation of angle between each vector of document with query vector where both are same type of vectors.
In vector space model it is easy to compute cosine of angle in between both vectors of query and documents rather we computer angle between them.

For example

$$\text{Cos}\theta = \frac{d_2 \cdot q}{\| d_2 \| * \| q \|}$$

In above dot product of document d_2 and query q, and $\|d_2\|$ is norm of document vector and as well as in similar way norm of query q is $\|q\|$. Vector norm is calculated by following formula.

$$\|q\| = \sqrt{\sum_{i=1}^{n} q^2 i}$$

if a cosine value is zeroing it means the document vector and query vector are orthogonal and have no match between them.

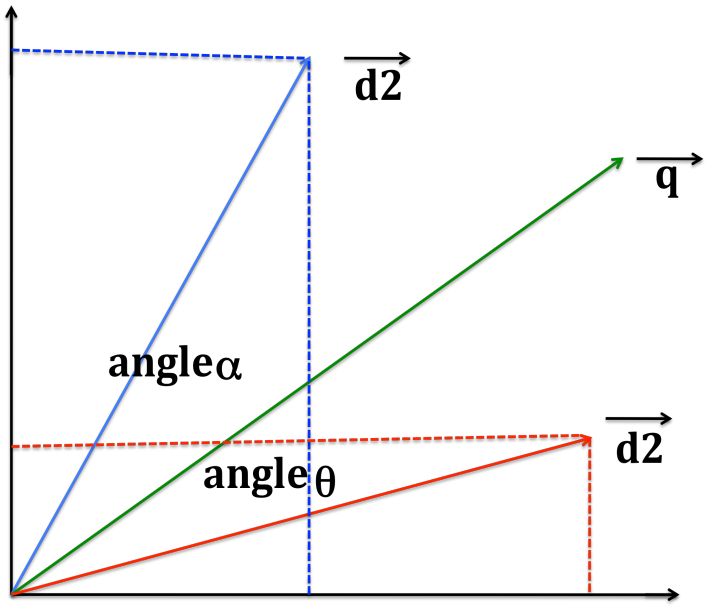

Figure 8. Vector space similarity [9].

In vector space model similarity angles between vectors in space easily determine the results of the model to non-technical experts. One can easily see the query and document vectors in space. Application of vector space model is in automatic text categorization and document clustering.

No of advantages of vector space model over Boolean model are:

- It is based on linear algebra.
- It allows similarity between documents and queries.
- Documents ranking as according to feasible relevance.
- Document and query term weighting are not binary.
- This model allows temporary matching.

4.4 Probabilistic Approach

Defining term weighting approach normally relies on probability theories.
Probability of any thing for example probability of relevance define as P (R), which is structure by using the experiment concept, these experiment is the way through which we made the observation. We take P (R) sample space should be {irrelevant, relevant} in which if we take the R as random variable values {0,1}, means 0= irrelevant value and 1 = relevant value [4].
For experiment purpose pick one document from the collection randomly. If we already know the total no of relevant document for example in collection 100 are relevant and total collection consist of one million documents (1 million) then probability can be define as:

Probability of relevance P (R = 1) = 100/1000000 = 0.0001

4.4.1 Probabilistic model of Indexing

In early 1960 Bill Maron and Larry Kuhn's (Maron and Kuhn's 1960) presented probabilistic indexing model. In this indexing model they did not particularly hit the automatic indexing, they preferred manual indexing, so during indexing which run for various Terms 'T' which apply to document 'D' given a probability P (T|D) to term T. for each document there is a set of possible index terms as:

$$\text{Weighted as} = P(T|D),$$

In which probability is P (T|D).

$$P(D|T) = P(T|D)\, P(D)/ P(T)$$

If we rank the documents by using P (D|T), so this is the probability by which document D is relevant. So document is ranked by P (T|D) P (D) this is value comparable with the value of P(D|T), if we see the value P(D) it is probability relevance of Document [4].

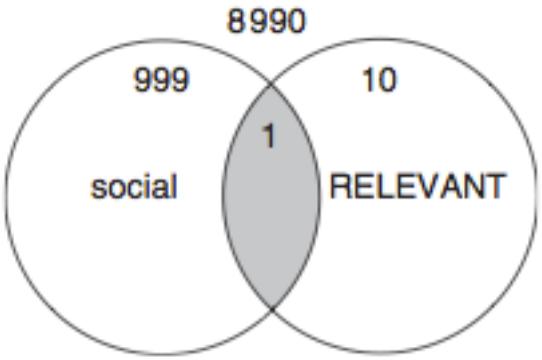

Figure 9. Collection given the query term political (Van diagram).

4.4.2 Probabilistic Model

Process of information retrieval is uncertain the basic criteria through which the information need is converted into query, document is converted into index terms, and query/document terms matches are computed are far from exact. A probabilistic information retrieval model tries to represent the probability of relevance of a document by a query. Actually it calculates the similarity coefficient between document and queries in which probability of document $\underline{\boldsymbol{d_j}}$ should be relevant to a user query called $\boldsymbol{q}$. if r represent the binary relevance judgment regarding the set of N documents D as with to q then coefficient similarity of the probabilistic model will be computed as:

$$SC\ (q,d_j) = P(r = 1\ |q,\ d_j)$$

As a document having increasing probability consider as top ranked and result in top of the list other documents will appear according to relevance probability of decreasing order.

The point that arises at this point is that the set of maximum relevant documents is unknown, and calculating the probability for them is very difficult task, so bootstrapping strategy which made up of in creating a initially probabilistic description regarding the ideal result set which represent the initial result set of documents, if we consider the probability of relevance is evaluated which based on the co-occurrence of terms in the query and documents from document set of reference. So iterative processing take place (it means strategy of using feedback of the user) which improve

the probability of result set, iterative strategy algorithm in terms of binary independence model is given in next section which describes iterative algorithm [4].

4.4.3 Binary Independence Model

Binary Independence retrieval model is classic probabilistic model of retrieving relevant documents in which documents and queries are shown as binary incidence vectors of term for example

$Dj = [w_{1j}, \ldots\ldots\ldots w_{Mj}]^T$

Where $wij = 1 \Leftrightarrow d_j$ consist of term t_i and $w_{ij} = 0$

Model considers term occurrences as independent, due to that "independence" is. More important consideration in the model is that the distribution of terms t_i in documents set R, which is relevant to query q, is totally changed from the set of non-relevant documents R, it means probability of occurring term t_i in document d_j changes based weather d_j is relevant or not relevant.

$$P(w_{ij} = 1 \mid d_j \in R, q) \neq p(w_{ij} = 1 \mid dj \in R', q)$$

So we got at the movement two probabilities one of the relevant documents probability and one of irrelevant document probability for term ti, because term ti occurs in many document casing these document as most relevant and absent in many non relevant document causing these non relevant but these document are also greater important in this model.

For example

$P(R|d_j)$ relevant in terms of relevant documents of term t_i.

$P(R'|d_j)$ irrelevant in terms of irrelevant documents of term t_i.

We can combine these two probabilities to define a query/document similarity coefficient. This is a ration between he probability that d_j is involved in the relevant set R and non-relevant set R'.

$$SC(d_j, q) = p(R|d_j)/P(R'|d_j)$$

4.4.4 Evaluation of the probabilistic Model

This model has advantages of ranking documents as to their decreasing probability of being relevant. User feedback as relevance feedback, which gives result of each query, is an advantage. But this depend on judgment of document relevance that is not easy to get accurate result and this model does not consider frequency of term occur in document and relies totally on the independent assumption of index term. Probabilistic model is less popular compare to VSM, which is state of the Art Information retrieval model.

Chapter 5: Knowledge Base Retrieval System

In this chapter we are going to describe the retrieval system and how it works, but the approach we are using in this process is based on reasoning where we gather knowledge and apply rules and get the appropriate answer from the facts as form of knowledge we have. But this IR system is based on probabilistic logic. This concept provides a framework for tasks to get required information from knowledge base.
The main function of this system is defined it to make it more easy for understanding. Further we will prove the implementation of retrieval of information for management decision and it will show the retrieval of data required by management. So information retrieval for information management is a mechanism, which allows modeling knowledge and reasoning [7].

5.1 What is Knowledge Base

A knowledge base term refers toward a database in which all the information are collected, shared, organized, searched and used. Knowledge bases are actually perform as a source for reader, for example Microsoft customer help or any customer representative helping customers using knowledge base, purpose of knowledge bases are to help answers to frequently asked questions [9]

5.1.1 Knowledge Base System

In knowledge base system where it reasons and uses knowledge base to solve complex problems. One common theme that gathered all knowledge-based systems to represent knowledge is called ontology and rules; it does not follow the conventional coding way of computer program.
In knowledge base system where it represent the facts about the world, like some shape of ontology. The early knowledge based system was initially for expert systems. Knowledge based system introduce toward the structure of the system, it display knowledge externally compare to the procedural coding system. One of first famous knowledge base was medical diagnosis system called Mycin. These knowledge base system use fact regarding the worlds as very simple claim as a simple and straight flat database and using rules to make reason about it.
We have no of advantages if we use rules to represent knowledge base

- **Maintenance and Acquisition**
- **Explanation**
- **Reasoning**

The one of the most recent advancement in the filed of knowledge based system is it has been adopted those system which use the Internet. As we know Internet use complex, unstructured data which does not fit to any data model. One of the model of knowledge based Internet system is known as semantic web [9].

5.2 Probabilistic Method

The term probabilistic can be defined as probabilistic characterization and reasoning tool by applying probability theory, for example Bayesian networks, hidden Markov. Likely also means belief that proposition of natural language shows the fact is true. If we see "beliefs" always have some numerical probabilities connected with them. So to measure the degree of justification of the belief we must need numerical probabilities, which calculate this justification. [14]
Probability distribution is the process, which can solve problems in probabilistic reasoning in expressive probabilistic logics. In decision-making science uncertainties evaluation are actually based on customer feedback evidence, which describe inferences. We have to ask in the evaluation of customer feedback uncertainties with in the knowledge.

So we can say probability theory can be used to do the reasoning process in the customer feedback retrieval. Bayesian networks got features that enable much aspect in probabilistic theory. So these features can be used in data analysis and management in the context of real world.

We take two proposition A and B, if we see numerical angle of belief walk behind the theory of probability.

1. Level of belief are taken as real numbers ranging from zero to one: $0 <= Pr(A) <= 1$.
2. Theory of A and B are not true at same time, because A and B are equally confined theories. Level of belief that any of them is true can be given as level of belief.
   $Pr(A \text{ or } B) = Pr(A) + Pr(B)$

If we see Bayesian networks it also describe the same as

$$P(X_i \mid A, Pa(X_i)) = P(X_i|Pa(X_i)).$$

$$P(B|A)\,P(A) = P(A,B) \qquad (1)$$

It follows by Bayes rule:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad (2)$$

If we consider numerical independence, Bayesian networks make subset of probabilities, which should not be derived from one another. Bayesian networks have very good inference system to improve probabilities, which can be obtained. So Bayesian network use to represent and manipulate uncertain knowledge. [14]

5.3 Representing Knowledge for Customer Feedback Retrieval

The representation of knowledge of Customer Feedback Retrieval for management decision formulized in a logic plan as in Datalog first order logic used.
To make a model of retrieval of customer feedback for decision-making a customer feedback is required. For this purpose we chooses a customer complain regarding TVM (ticket vending machine) at stations. One of the stations TVM (ticket vending machine) has some problem due to that customer feedback about that TVM.

Facts of problems:

1. Problem (tvm90, “coin stuck every time in ticket machine”) => problem 1.
2. Problem (tvm91, “ticket machine display problem”) => problem 2.
3. Problem (tvm92, “ticket machine taking money but not issuing ticket”) => 3.
4. Problem (tvm93, “ticket machine oyster reader not reading oyster card) => 4.

5. TVM (tvm90, “limehouse”) => 5.

6. TVM (tvm91, “Bow church”) =>6.

The main question here is to find the all stations with their TVM (ticket vending machine names.
In Datalog the facts related to cause of problem and their rule represented here in figure below [14].

1. Problems are declared as “problem” predicate with arguments of tvm, En and cause of problem Cn.
2. Stations making problem declared as ‘TVM’ predicate along with arguments of TVM, En, and station name Cn.
3. Rule of result uses a repetitive search strategy.
4. Result (T, S): = problem (T, P) & TVM (T, S);

```
/* Facts of problem */

problem (tvm90, “coin stuck every time in ticket machine”);
problem (tvm91, “ticket machine display problem”);
problem (tvm92, “ticket machine taking money but not issuing ticket”);
problem (tvm93, “ticket machine oyster reader not reading oyster card);

/* Facts of TVM stations */

TVM (tvm90, “limehouse”);
TVM (tvm91, “Bow church”);

/*rule*/

Result (T, S): = problem (T, P) & TVM (T, S);
```

Figure 10. Knowledge Representation For Customer Feedback.

5.4 Modelling Knowledge Base Retrieval Scenarios

Complex information retrieval job needed the modeling of information retrieval models. Using facts, and rules and querying which describes these models and solve the given

task in these models. Modeling information retrieval for management decision making (IR for MDM) using PDatalog.

Consider the following probabilistic Datalog programs.

5.3.1 TVM (Ticket vending machine):

This PD programme based on information given by customer feedback during ticket vending machine usage at different stations, customer information operating company of rail service via email regarding problems they facing during buying ticket vending machine.

Queries that need to be solved using probabilistic Data log.

- List of customers complain about TVM at stations.
- List of stations most problems about TVM complain.
- Which ID of TVM most problem at stations with station names.
- What kind of problem customer complains about TVM?

#TVM indexing

Customer (peter);

customer(palaf);

customer(paul);

customer(ben);

customer(charles);

customer(jun);

stations

station(station,limehouse);

station(station,stratfor);

```
station(station, greenwich);

station(sation, westferry);

station(station, poplar);

#customers

term(customer, peter);

term(customer, palaf);

term(customer, paul);

term(customer, ben);

term(customer, charles);

c_s(limehouse, peter);

c_s(stratford, palaf);

c_s(westferry, ben);

c_s(poplar, charles);

c_s(greenwich, jun);

#Ticket Vanding Machines

tvm(tvm90, limehouse);

tvm(tvm91, limehouse);

tvm(tvm92, westferry);

tvm(tvm93, westferry);

tvm(tvm94, poplar);

tvm(tvm95, stratford);

tvm(tvm96, stratford);

tvm(tvm90, greenwich);

tvm(tvm92, devensroad);

# list of customer complaining
```

_echo("Customer Complains");

retriev_c Customer(C):- customer(C);

query customers

?- retriev_c(C);

# query customers stations
_echo("Customer complain stations");

retriev_s Cust(C,S) :- retriev_c(C) & c_s(S,C);

5.3.2 Hazards

This ontology based on information given by customer feedback during rails system usage during their travel using the stations and train services at different stations, customer complaining about hazards to operating company of rail service via email regarding injuries and hazard he realizing during travel and using stations services.

- List Type of Hazards at different station on rail services.
- List of stations most problems about Hazard complain.
- List service at stations causes most problems Hazard with station names.
- List what kind of Injuries causes by Hazard to customer.

#Hazard complaint

customer(1, danil);

customer(2, peter);

customer(3, dominic);

customer(4, jack);

customer(5, ravi);

customer(6, umer);

customer(7, parker);

customer(8, lamen);

```
#Customer type

profession(1,"it engineer");

profession(2,student);

profession(3,doctor);
profession(4,"civil servent");

profession(5,"sale man");

profession(6, jobless);

profession(7, "stock taker");

profession(8, lecturer);


#Hazard types

hazard(1,slip);

hazard(2,trip);

hazard(3, "stuck in door");

hazard(4, "slippery surface");

hazard(5, "escalator slippery");

hazard(6, "platform edge narrow");

hazard(7, "rain water on platform");

hazard(8, "stuck in lift");

hazard(9, "lift very bad smell");

hazard(10, "station concourse slippery");

hazard(11, "train door close very quickly");

hazard(12, "train floor slippery");

hazard(13, "luggage unattended");

hazard(14, "too many people at platform");
```

```
#stations

station(1,"bow church");

station(2,"stratford");
station(3,"pudding mill lane");

station (4, "devons road");

station(5, "langon park");

#stations

station(1,"bow church");

station(2,"stratford");

station(3,"pudding mill lane");

station (4, "devons road");
station(5, "langon park");

station(6, "Al saints");

station(7, "poplar");

station(8, "west ferry");

station(9, "lime house");

station(10, "shadwell");

station(11, bank);

station(12, "canary wharf");

station(13, "Heron Quay");

#hazardious stations customer wise

hazard_station(1,1,"bow church");

hazard_station(2,2,"stratford");

hazard_station(3,3,"pudding mill lane");

hazard_station(1,4,"bow church");

hazard_station(4,5,"langdon park");
```

```
hazard_station(3,12,"canary wharf");

hazard_station(6,13,"Heron Quay");

#customer complaint

complaint(1,1,10);

complaint(2,1,13);

complaint(3,2,2);

complaint(1,4,5);

complaint(6,7,2);

complaint(2,8,9);

complaint(5,13,8);

complaint(8,11,5);

#Queries

_echo("list of hazards at different stations");

stations Station(Sq,St) :- station(Sq,St);

h_station Hazard(Hsq, Hst):- hazard(Hsq,Hst);

?-  stations(Sq,St);

?-  h_station(A,B);

_echo("Stations with most common hazard");

retrieve(S,H):- station(ID,S) & hazard(ID,H);

?- retrieve(S,H);

_echo("Station most complain by customer");

cust (CID,C):- customer(CID,C);
stat (CID,SID,S):- station(CID,SID,S);
comp (C,S,Comp):-  comp(C,S,Comp);
match(C,S):- cust(CID,C) & stat(CID,SID,S);
?- match(C,S);
```

```
# This Query bring All most complains stations with  Customer who
# compalins  with their profession.

_echo("Station most complain by Customers with profession:");

cust (Cid,C):- customer(Cid,C);

prof(Cid,P):-  profession(Cid,P);

stat(Sid,S):-  station(Sid,S);

match(Cid,C,P):-   cust(Cid,C) &  prof(Cid,P);

h_s_c(C,P,S):-   hazard_station(Cid,ID,S) & match(Cid,C,P);

?- h_s_c(C,P,S);


# This query bring all customer wise complaint with station name and
# their complain type.

_echo("Station wise complain type");

 t_h(Hid,H) :-  hazard(Hid,H);

 s_h(Sid,S):-   station(Sid,S);

 c_h_c(Cid,C):- customer(Cid,C);

 c_h(Hid,H,Cid,Sid):-     t_h(Hid,H) & complaint(Cid,Sid,Hid);

 c_s(Sid,S,Hid,Cid):-     s_h(Sid,S) & complaint(Cid,Sid,Hid);

 c_c_h_c(Cid,C,Hid,Sid):- c_h_c(Cid,C) & complaint(Cid,Sid,Hid);

 h_c_c_w_s(C,S,H):- c_h(Hid,H,Cid,Sid) & c_s(Sid,S,Hid,Cid) &
c_c_h_c(Cid,C,Hid,Sid);

?- h_c_c_w_s(C,S,H);
```

5.3.3 Services

This ontology based on information given by customer feedback during their travel using the stations and train services. Customer complaining about hazards to operating company of rail service via email regarding injuries happened?

- List All Injuries of Customers.
- Inform manager serious Injuries

```
#customer injuries

customer (1,marry);

customer (2,margarate);

customer(3,anthony);

customer(4,pavan);

customer(5,jun);

customer(6,farooq);

customer(7,shahid);

customer(8,elmar);

customer(9,jack);

# injuries at stations
injury(1,"slip on stairs");

injury(2,"nail cut in train door");

injury(3,"sliped at platform ");

injury(4,"stuck in door ");

injury(5,"tvm machine cut and ");

injury(6,"push by customer");

injury(7,"hit by train");

injury(8,"contious at platform");

injury(9,"door hit hand with force");

#serious injury

serious_injury(1,7);
```

```
serious_injury(2,2);
serious_injury(3,3);
#service
#inform manager
manager_area(lucy,"DLR");
manager_area(smith, "Underground");
manager_area(tomm, "Overground");
#customer complaint with injury type
cust_inj_comp(1,1,4);
cust_inj_comp(2,2,5);
cust_inj_comp(3,3,3);
cust_inj_comp(4,4,1);
cust_inj_comp(5,5,2);
cust_inj_comp(6,6,6);
cust_inj_comp(7,7,9);
cust_inj_comp(8,8,7);
cust_inj_comp(9,9,8);

#list all injuries of customers:

_echo(" Customer Complain List");
cust(Cid,C) :- customer(Cid,C) & cust_inj_comp(CmId,Cid,Iid);
match(Iid,I) :- injury(Iid,I) & cust_inj_comp(CmId,Cid,Iid);
retrieve(Cmid,C,I) :- cust(Cid,C) & match(Iid,I) &
cust_inj_comp(Cmid,Cid,Iid);
```

```
_sort(retrieve);

?- retrieve(Cmid,C,I);

#inform manager serious injury :

inform_manager_evidence(Name,I) :-
        cust_inj_comp(Id,Cid,Iid) &
        serious_injury(Sid,Id) &
        manager_area(Name,Dept) &
       injury(Iid,I);

?-  inform_manager_evidence(Name,I)
```

Chapter 6: Evaluation and Results

6.1 Evaluation of performance of Customer Feedback Retrieval System

The information Retrieval system generates simple and complex queries results. The criteria in simple queries is simple and criteria in complex queries is bit complex to get the answer of the query so both simple and complex queries result are compare in this method to check the compatibility of the queries.

Method

6.1.1 Participant

The study involves the comparison of the queries with different systems (SQL) to check validity, compatibility and testability of the queries.

6.1.2 Material

Standard SQL-92 is used for testing the compatibility, performance, correctness, soundness of the queries with already run queries on Datalog Engine (HySpirit Probabilistic Datalog (PD) Engine).

6.1.3 Procedure

There are around 12 queries run on Datalog Engine in which some queries are simple some medium and some are complex but all the queries bringing result according to the user requirement as there are no difference found between Datalog Engine queries and SQL queries result set. As we run same queries on SQL engine and we got the same result as we have found result on running the queries on Datalog Engine so this proved that all the queries run on Datalog engine are 100% correct in compatibility and

performance as all queries run fast and accurate as no too much time has been taken by queries run on Datalog engine.

6.1.4 Comparison of Queries

Here are some examples of query comparison of SQL and Datalog with working implementation, for comparison.

Example 1: Retrieve all tuples /rows from table/facts:

SQL:

SELECT * FROM TVM;

SQL bring all the TVMs from TVM table.

DataLog:

? - TVM (T, S); // browse the tvm's entire name, stations where it is located.

Example 2: Restriction

SQL:

SELECT * FROM tvm WHERE station = "stratford"

Datalog:

?- station(T,Tid,Sid,S);

Example 3: Restriction with Project, Unique record.

SQL:

SELET DISTINCT complaint_no, customer_name, complaint_station, complaint_*desc FROM* complaint
GROUP BY complaint_no,customer_name,complaint_station,complaint_desc
WHERE station_name = "CANARY WHARF";

Bring all DISTINCT record of particular customer complains.

Datalog:

User complaint (A, B, C): = complaint (A, B, C, Station);

? - User complaint (A, B, C);

Table 4. List of Queries (Datalog Engine)

Query	**Description**
Query 1	List of customer complaining
Query 2	List of customer complaining stations
Query 3	List of most problem TVM stations
Query 4	List of problems in TVM (ticket vending machine)
Query 5	All injuries of customers
Query 6	Inform manager serious injuries at stations
Query 7	Types of hazard at different stations
Query 8	Most common stations for hazard
Query 9	Stations most complain by customers with profession
Query 10	Customer wise complaint with stations name and complain type
Query 11	Customer service complaint stations
Query 12	Customer service complaint and stations names

Table 5. Query Result Comparison Ratio

Datalog Query	**SQL Query**	**Comparison Ratio**
Query 1	SQL Query 1	100%
Query 2	SQL Query 2	100%
Query 3	SQL Query 3	100%
Query 4	SQL Query 4	100%
Query 5	SQL Query 5	100%
Query 6	SQL Query 6	100%
Query 7	SQL Query 7	100%
Query 8	SQL Query 8	100%
Query 9	SQL Query 9	100%
Query 10	SQL Query 10	100%
Query 11	SQL Query 11	100%
Query 12	SQL Query 12	100%

6.1.4 Results

The comparison of Datalog queries with standard SQL query shows that the required result from Datalog queries is quite similar with the result obtain from standard SQL queries and this result is nearly 100%. Speed of data retrieval in Datalog queries is quite fast and reliable and rules are efficiently retrieving required result from the facts.

6.1.5 Conclusion

In this Information retrieval system where we are extracting the information for management for decision-making is quite reliable and can extract all the required information smoothly and present required data to management to take decision on the basis of customer feed backs.

6.2 Testing and Validation of Customer Feedback Retrieval System

The information retrieval system is done by using the knowledge base where we produces the list of facts and then we gathered these facts using rules and then we applied queries to obtain the required relevant information. These queries are tested using Datalog Engine where each query provides exact result as required by the management.

6.2.1 Experiment

Experiment start by first writing facts these facts are extracted from the feedback of the customers regarding hazards and services at different stations during rail journey to and from home and then rules are made to get these facts data out using queries applied on rules, so this is step by step process in which first we create facts then apply rule and then run the query as below.

Table 6. List of Facts

Customer (1,”denial”)	Profession (1, “doctor”)	Hazard (1, “slip”)
Customer (2, “peter”)	Profession (2, “engineer”)	Hazard (2, “trip”
Customer (3,dominic)	Profession (3, “jobless”)	Hazard (3, “stuck in door”)
Customer (4, ”jack”)	Profession (4, “teacher”)	Hazard (4,”slippery surface”)
Customer (5, “Ravi”)	Profession (5, “IT”)	Hazard (5,”escalator slippery”)
Customer (6, “umer”)	Profession (6,”civil servant”)	Hazard (6,”platform edge narrow”)
Customer (7, “parker”)	Profession (7, “sales man”)	Hazard (7,”water at platform”)
Customer (8, “lamen”)	Profession (8,”stock taker”)	Hazard (8,”luggague unattended”
Customer (9, “thapa”)	Profession (9, “lecturer”	Hazard (9,”concourse Slippery”)
Customer (10, “Robert”)	Profession (10,student)	Hazard(10,”stuck in lift”)

Table 6. Table of Rules and Queries

_echo ("Station most complain by Customers with profession:");
 	
cust (Cid,C) :- customer(Cid,C);
 	
prof(Cid,P) :- profession(Cid,P);
stat(Sid,S) :- station(Sid,S);
 	
match(Cid,C,P):- cust(Cid,C) & prof(Cid,P);
 	
h_s_c(C,P,S) :- hazard_station(Cid,ID,S) & match(Cid,C,P);
 	
?- h_s_c(C,P,S);

_echo ("Station wise complain type");
 	
t_h(Hid,H) :- hazard(Hid,H);
 	
s_h(Sid,S) :- station(Sid,S);
 	
c_h_c(Cid,C) :- customer(Cid,C);
 	
c_h(Hid,H,Cid,Sid) :- t_h(Hid,H) & complaint(Cid,Sid,Hid);
 	
c_s(Sid,S,Hid,Cid) :- s_h(Sid,S) & complaint(Cid,Sid,Hid);
 	
c_c_h_c(Cid,C,Hid,Sid) :- c_h_c(Cid,C) & complaint(Cid,Sid,Hid);
 	
h_c_c_w_s(C,S,H) :- c_h(Hid,H,Cid,Sid) & c_s(Sid,S,Hid,Cid) &
 	
c_c_h_c(Cid,C,Hid,Sid);
 	
?- h_c_c_w_s(C,S,H);

Table 7. Query Result.

Station most complain by Customers with profession: # ?- h_s_c(C,P,S) # ?- PROJECT ALL[$1,$2,$3](h_s_c) (danil,"it engineer","bow church") (peter,student,stratford) (dominic,doctor,"pudding mill lane") (danil,"it engineer","bow church") (jack,"civil servent","langdon park") (dominic,doctor,"canary wharf") (umer,jobless,"Heron Quay") #7 tuples
Station wise complain type # ?- h_c_c_w_s(C,S,H) # ?- PROJECT ALL[$1,$2,$3](h_c_c_w_s) (dominic,stratford,trip) (umer,poplar,trip) (danil,"devons road","esclator slippery") (lamen,bank,"esclator slippery") (ravi,"Heron Quay","stuck in lift") (peter,"west ferry","lift very bad smell") (danil,"bow church","station concourse slippery") (peter,"bow church","luggage unattended") # 8tuples

7.3.2 Results

The above experiment (Table 7) validated the correctness of the information retrieval for management decision system. The out put query result in result table shows that required information has been extracted from the customer feedback.

Chapter 7: Conclusion and Further Work

7.1 Further Work

During this project evolution, some points appear beyond its main objective in very short time. In this section we will discuss ways to improvement and enhancement of the system. Currently this information retrieval system for management decision making provide very basic approach showing how we can extract the required relevant information from user feedback data. There is always a space for enhancement as requirement is changing unlimitedly. The result of this system is as full fill the requirement of the management decision need but we can make this system more efficient for larger data set and even we can make this system more efficient for different type of documents.

7.2 Conclusion

This information retrieval system for management decision provide many significant benefits for management because using this management have more correct information extraction from customer feed back data on the basis of these information retrieval management can get reliable and accurate data to make decision in future regarding enhancement of services at different stations as well as reducing the level of hazards customer facing at different stations during travel on rail service.
Along with this management can use this system to make decision of planning about spending money on stations. So on the basis of retrieval result set management can decided where there is need to spend money to improve services and stations.

Reference:

1. Manning, Christopher D (2008) An Introduction to Information Retrieval, [http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf]

2. A Study on Semantic similarity and its application to clustering Enabling the classification of textual data by: Dr Montserrat Batet

3. Baeza-Yates, R. (Ricard) (1999) Modern Information Retrieval Online Edition Cambridge university Press (2009) [http://nlp.stanford.edu/IR-book/pdf/01bool.pdf]

4. Ays ¸e Go ¨ker, John Davies (2009: Information Retrieval: Searching in the 21st Century

5. Stefano Ceri,marco, Piero,Silva, Emanual, Alessandro: Web Information Retrieval.

6. Trip Advisor [http://www.tripadvisor.co.uk/Attraction_Review-g186338-d246022-Reviews-Docklands_Light_Railway-London_England.htm]

7. Miguel Martinez-Alvarez and Thomas roelleke Modelling Probabilistic Inference Networks and Classification in Probabilistic Datalog

8.Yuanbo Guo, Zhengxiang Pan, and Jeff Heflin An Evaluation of Knowledge Base Systems for Large OWL Datasets

9. Wikipedia The Free Encyclopedia.

10.Thomas Roelleke · Hengzhi Wu · Jun Wang · Hany Azzam Modelling Retrieval Models in a Probabilistic Relational Algebra with a new Operator: The Relational Bayes

11. On the Modelling of Ranking Algorithms in Probabilistic Datalog Thomas Roelleke and Marco Bonzanini and Miguel Martinez-Alvarez

12. Measuring Search Effectiveness
[https://www.creighton.edu/fileadmin/user/HSL/docs/ref/Searching_-Recall_Precision.pdf]

13. Precision and Recall of Five Search Engines for Retrieval of Scholarly Information in the Field of Biotechnology
[http://www.webology.org/2005/v2n2/a12.html]

14. Probabilistic knowledge Base System from Forensic Evidence Analysis
[http://www.jatit.org/volumes/Vol59No3/21Vol59No3.pdf]

15. Expert System
[http://www.cieer.org/mthomas/thesis/chapter2.html]

16. Expert System
[http://shodhganga.inflibnet.ac.in/bitstream/10603/7654/6/06_chapter%202.pdf]