**SYNOPSIS**

Nowadays, with rapid growing of opinionated text on the Web, mining aspect-level opinions has become a promising means for online public opinion analysis. In particular, the booming of various types of online media provide diverse yet complementary information, bringing unprecedented opportunities for public opinion analysis across different populations. Along this line, information analysis algorithm (IAA) a novel topic model for complementary aspect-based opinion mining across asymmetric collections.

We gathered data from online shopping site. In our Example, we took Amazon reviews based on electronics.Unfortunately, very few studies has been done for sentiment analysis in Persian online shopping and the existing works have many limitations in their performances in differentiating between a weak and a strong mood in sentimental sentences. This opinion contains lots of information which can be useful for future systems.

Opinion mining is an emerging field which combines lots of tasks to find the correct Opinion Words and Targets which describes the product or express the sentiment towards product. A new lexicon based opinion mining method for Persian online shopping which considers the effect of intensifier adjectives in extracting the exact opinion of a review. These opinions are very valuable in decision making process. Analyzing and extracting the actual opinion throughout these reviews manually is practically difficult since there are large numbers of reviews available in the various aspects.

**INTRODUCTION**

Opinions are plays very important role in today’s life for purchasing any item or product. This opinion contains lots of information which can be useful for future systems. Opinion mining is an emerging field which combines lots of tasks to find the correct Opinion Words and Targets which describes the product or express the sentiment towards product. Main focus is on Topical Relation, which shows the relation between opinion target and opinion words. These topical relations are useful to gain the idea of which topics are mostly discussed by customers for particular product.

Opinion is a statement which expresses sentiment towards some product, service, movies or even sentiment about a particular person. In today’s world, taking opinion from others is rapidly increasing in every other field. For example, if someone wants to buy a laptop from online site, that person will first reads reviews which are submitted by other customers. If that review contains the positive comments, then only other customers are likely to buy that product. Like this opinion or reviews given by other users are play important role in purchasing a product, buying a car and in marketing field.

The data set which requires data that changes over time must use incremental approach, which allows past updating of past results using new data instances. Another challenge is to tackle grammatical errors. We can improve our results by mapping theses grammatical errors to correct words. As we are dealing with online reviews, there are chances that the data can be noisy, unstructured and dynamic in nature. Removal of such noisy data is also a challenging task. Our main work is to find out the Opinion Targets and Words.

Opinion Targets are objects on which user express their Opinion or Expressions. Opinion words are those adjective words which are used to express the sentiment about that object. To extract opinion targets we made data set of different objects which can mostly use by the users. For example, in our work we are taking electronic reviews, so most of the opinion targets will be (Price, Display, Screen size, Battery, Performance etc.) Another kind of methods is distributional similarity which is popular in text classification. Those methods are assumed that context words of features belonged to same aspect are similar. Some similarity measures, such as Euclidean distance, Cosine, Jaccard, etc., are used to calculate the similarity of documents. Some other work also computes the Pointwise Mutual Information (PMI) between words and then clusters words with K-Means based on cosine similarity of PMI weight. The work in uses two constraints based on sharing words and lexical similarity to generate labeled data.

**SYSTEM ANALYSIS**

**EXISTING SYSTEM**

Opinion is a statement which expresses sentiment towards some product, service, movies or even sentiment about a particular person. In today’s world, taking opinion from others is rapidly increasing in every other field. For example, if someone wants to buy a laptop from online site, that person will first reads reviews which are submitted by other customers. If that review contains the positive comments then only other customers are likely to buy that product.

The existing system for user access information Unstructured of data in the process so user collect didn’t clear information process this problem used to performance slow. Like this opinion or reviews given by other users are play important role in purchasing a product, buying a car and in marketing field. This relation shows relativity between opinion words and opinion target. The topics are nothing but the features or the attributes of the object, on which customers are likely to express their opinion. In this paper we represent, how we are dealing with the topical relation in our work and how this topical relations are useful in future.

**DISADVANTAGES**

* The mining process used to information are not identity of the user access in the problem
* Thus performance based low of the opinion mining problem
* Unstructured of data didn’t clear data in the process
* Maintenance is very difficult of a mining process
* Unsecured of information collecting process

**PROPOSED SYSTEM**

These are called opinions which are valuable in the decision-making process. Therefore, the World Wide Web has become a huge repository containing different kind of opinions and thoughts of the people. However, to get benefits from these accumulated opinions, the contents should be extracted in to features such as “foods”, “services”, “environment” in the restaurant domain and analyzed properly since those opinions are written in complex sentences and in the row text format. Information analysis algorithm (IAA) so this technique used to all users easy to access of the data and its easy maintenance of a process As previously mentioned we utilize the favor of intensifiers in our work and we also detect adverbs and adjective intensifiers in the sentences.To do that, we need two dictionaries including adjectives and adverbs each with polarity labels. The importance of using the intensifiers is the difference in the amount of satisfaction or emotional load that is transmitted to the reader. For example, the sensational mood sentences "the battery is a bit weak" and "the battery is very weak” is obviously not the same so it should be considered in the models.

**ADVANTAGES**

* The opinion mining process used to easy to access online data
* Searching speed up data processing so user have a fast access of information
* Maintenance very easy of a data processing to mining
* Thus data are classification of user to access in the process
* Secured data access and it opinion mining process of a model
* The performance are high of a process in the mining to data

**REQUIREMENT ANALYSIS**

**HARDWARE REQUIREMENTS**

PROCESSOR : Core2duo

CLOCK SPEED : 3.0 GHZ

RAM SIZE : 2 GB

HARD DISK CAPACITY : 250 GB

MONITOR TYPE : 15 INCH COLOR MONITOR

KEYBOARD TYPE : INTERNET KEYBOARD

**SOFTWARE REQUIREMENTS**

Operating System : WINDOWS XP PROFESSIONAL

Front end : DOTNET

Back end : WEEK A JAR

Documentation : Ms-Office

**2. LITERATURE SURVEY**

**2.1 STATISTICAL ANALYSIS AND MODELLINGOF SPATIAL POINT PATTERNS**

**AUTHORS:JANINE ILLIAN**

The analysis of spatial pattern data usually approached? a simple graphical representation of the pattern of objects as a point map is a very useful preliminary step towards understanding its properties: visual inspection provides a qualitative characterization of the type of the pattern even if rather vague terms are used in the initial description (clustered, aggregated, clumped, patchy, regular, inhibited, uniform, even); it may also indicate correlations among marks or between the point density and spatial covariates, i.e. other random structures which influence the point distribution such as soil property or physical influences. However, for a precise quantification and a more standardized description and finer distinction between types of spatial behavior, appropriate statistical methodology has to be applied. These methods provide information on more subtle differences in spatial structures that are not apparent to the naked eye.

Many different aspects of the nature of a specific spatial point pattern may be described using the appropriate statistical methods. The simplest of these is the point intensity, i.e. the average number of objects per unit area or volume, if point density can be considered constant across space. Note that this resembles the use of the sample mean x in classical statistics. If the point density is variable, intensity maps may be constructed and these may be related to maps describing the values of covariates across space, perhaps in the context of geographical information systems. The analysis of marked point patterns is more interesting and often provides deeper insight into the processes that are causing the pattern than an analysis of unmarked point patterns. Often the marks are ‘qualitative’, i.e. the pattern is multivariate and consists of several types of points, e.g. different species, ages and size classes. This kind of point pattern data may be regarded either as a superposition or union of single-type point patterns, or alternatively as a single point pattern with different types of labels where the labels indicate the type of points. Even more general information may be assigned to the individual points. The marks can be continuous variates, vectors of variates or even stochastic processes. Examples from forestry include the diameter at breast height (dbh) of a tree, last year’s growth increment, and a time series of dbh over several consecutive years. Other examples are particle diameter, size, and shape. a common feature of these measurements is that they describe the status or property of the object associated with a point location. Data sets with ‘quantitative’ (or real-valued) marksarehighly complex as they reflect various correlations among the objects represented by the points and contain an abundance of information on the system of objects. Point process statistics may be used to detect these correlations and hence provide information contained in these data sets.

**2.1.1 WATERSTRIDERS**

The application of point process statistics in biology is not restricted to nonmotileorganisms such as plants; the methods are also suitable for cross-sectional data on positions of animals at a specific time point. Which is based on a photograph, shows the positions of Pala arctic water striders (limnoporusrufoscutellatus) on a water surface? Water striders are arthropods that live on theater surface and move at high speed from time to time. Individuals communicate by sending signals along the water surface by vibrating their front legs. The spatial patterns formed by individuals contain information on the animals’ behavior. Various aspects of behavior, such as habitat selection at different stages in the lifecycle (larval stages, juvenile, and adult), territoriality and cannibalism, are of ecological interest and have been studied in the literature, both in experiments and in the natural environment. The figure shows the last larval stage (stage 5) in a subrectangleof a water surface of irregular shape. The rectangle was chosen such that methods for stationary point processes can be applied even though the pattern is very small, i.e. the pattern is treated as if it were a small rectangular area in a very large water surface with many water striders. Modern statisticians would probably prefer to observe the entire finite pattern and apply the methods of chapter 3 for finite point processes.

**2.1.2 Sampling and data collection**

Data collection methods are strongly dependent on the objects represented by the points, the objectives of the study and the available resources, and all these are different for different applications. Generally, the best choice is only the best choice for a very specific example, due to the specific situation and study aims. However, some general guidelines for data collection can be given. Central aspects that need to be considered include the spatial scale, the relationship to the environment, the morphology, size and density of the objects, and available methods. From a statistical point of view, the aim of data collection methods is to optimiseunbiasedness and representativeness and to control sampling errors. The two main aspects in a strategy for data collection are appropriate sampling and appropriate measurement methods as these are both sources of uncertainty. Sampling concerns the strategy of extracting information on point patterns in a reallifesituation, whereas measurement is the technical realization of the data collection approach. Sampling is a traditional area of statistics and many useful methods have been developed in the literature. Recently, measurement technologies have Undergone rapid development such that more and more extensive and high-quality point pattern data have become available.

**2.1.3 Field Methods**

Once the window has been chosen, data can be collected. Clearly, the choice of the data collection method has a pronounced impact on the choice of the statistical analysis methods. The most informative type of point pattern data consists of the cartesian coordinates of all points. Often their measurement is too laborious, or partial or condensed information might be sufficient for a particular study. For example, measurements are sometimes taken in the field and are directlysummarised in indices that describe the spatial behavior of the pattern.

**2.1.4 Marked Point Processes**

Marked point processes are generalizations of point processes and are highly relevant in practical applications. each point xi is assigned a further quantity mxi,which provides additional information on the object represented by the point, as discussed in the examples in (qualitative marks, cell types) and (quantitative marks, particle diameters). in this book the mix are usually integers or real numbers, but much more general marks may also be considered. A marked point process is denoted by m.

**2.2 DISCOVERING CO-LOCATION PATTERNS FROM SPATIAL**

**DATASETS: A GENERAL APPROACH**

**AUTHORS:YAN HUANG, SHASHI SHEKHARAND HUI XIONG**

Given a collection of Boolean spatial features, the co-location pattern discovery process ends the subsets of features frequently located together. For example, the analysis of an ecology dataset may reveal symbiotic species. The spatial co-location rule problem is different from the association rule problem since there is no natural notion of transactions in spatial data sets which are embedded in continuous geographic space. in this paper, we provide transaction-free approach to mine co-location patterns by using the concept of proximity neighborhood. a new interest measure, a participation index, is also proposed for spatial co-location patterns. the participation index issued as the measure of prevalence of a co-location for two reasons. first, this measure is closely related to the cross function, which is often used as a statistical measure of interaction among pairs of spatial features. Second, it also possesses an anti-monotone property, which can be exploited for computational efficiency. Furthermore, we design an algorithm to discover co-location patterns. This algorithm includes a novel multi-resolution pruning technique. Finally, experimental results are provided to show the strength of the algorithm and design decisions related to performance tuning. co-location patterns represent subsets of Boolean spatial features whose instances are often located in close geographic proximity Boolean spatial features describe the presence or absence of geographic object types at different locations in a two dimensional or three dimensional metric space, such as the surface of the earth.

**2.2.1 Geometric Approach:**

The geometric approach can be implemented by neighborhood relationship-based spatial joins of table instances of prevalent co-locations of size with table instance sets of prevalent co-locations of size In practice, spatial join operations are divided into a alter step and a Retnementstep to efficiently process complex spatial data types such as point collections in a row instance. In the alter step, the spatial objects are represented by simpler approximations such as the mbr - minimum bounding rectangle. There are several well-known algorithms, such as plane sweep space partition and tree matching which can then be used for computing the spatial join of mbrs using the overlap relationship; the answers from this test form the candidate solution set. in the tenement step, the exact geometry of each element from the candidate set and the exact spatial predicates are examined along with the combinatorial predicate to obtain the final result.

Observation was that some properties of proximity neighborhood cliques obey the downward inclusion property necessary for apriority-based rule mining. The cardinality of table instances does not obey this property but the proposed participation index does, allowing interactive pruning. in addition, the participation index has a spatial statistical interpretation as an upper-bound on the cross-function, a classical spatial statistical measure of association for binary.

The co-location miner, an algorithm for mining co-location patterns, was presented and analyzed for correctness, completeness and computation cost. Design decisions in the proposed algorithm were evaluated using theoretical and experimental methods. Empirical evaluation shows that the geometric strategy performs much better than the combinatorial strategy when generating size-2 co-locations; however, it becomes slower when generating co-locations with more than 2 features. the hybrid strategy integrates the best features of the above two approaches. furthermore, when the locations of the features tend to be spatially clustered, which is often true for spatial data due to spatial-autocorrelation, the computation cost of the co-location miner can be significantly reduced with a multi-resolution alter.

Also considered only Boolean features here. In the real world, the features can be categorical and continuous. There is a need to extend the co-location mining framework to handle continuous features. Third, we plan to evaluate the impact of multi-resolution altering on overall performance of the proposed algorithms using real world datasets which exhibit strong spatial auto-correlation. Finally, if locations of features change over time, it is possible for us to identify some spatio-temporal association patterns.

**2.3 ANALYSIS OF SPATIAL POINT PATTERNS USINGBUNDLES OF PRODUCT DENSITYLISA FUNCTIONS**

**AUTHORS:NOEL CRESSIEA ND LINDA BRANTC OLLINS**

The analysis o f a spatial point pattern i s often involved with looking or spatial structure, such as clustering or regularity in the points (or events). For example, it is of biological interest to characterize the pattern of tree locations in a forest. This has traditionally been done using global summaries, such as the k-function or its differential, the product density function. in this article, we define a local version of the product density function for each event, derived under a definition of a local indicator of spatial association (Lisa). This product density lisa functions can then be grouped into bundles of similar functions using multi variatehierarchicalclusteringtechniques. the bundles can then be visualized by a repotting of the data, obtained via classical multidimensional scaling of the statistical distances between functions. thus, we propose a different way of looking for structure based on how an event relates to nearby events. we apply this method to a point pattern of pine saplings in a finish forest and show remarkable here to forerun discovered, spatial structure in the data.

**2.3.1 EDGE EFFECTS**

Thus far, we have ignored edge effects in the development of a product density Lisa function. Clearly, for xi observed near the border of the observation region a, the product density lisa function will underestimate him contribution fix to the global product density function. to investigate the effect of edges on the distribution of the product density lisa function, consider its expected value for a homogeneous Poisson process with intensity.

Application of the exploratory method described in this article is straight forward. However, the reader should keep in mind the following caveats and areas for future work on this topic. first,e are using hierarchical clustering and classical multidimensional scaling developed for independent multivariate observations on spatial data that are dependent, with some adjustment for the dependence through the use of statistical distances. Second, to do that adjustment we are using covariance calculated assuming a homogeneous Poisson process when the point patterns may exhibit clustering or regularity.

**2.4 A JOIN-LESS APPROACH FOR CO-LOCATION PATTERN MINING: A SUMMARY OF RESULTS**

**AUTHORS:JIN SOUNG YOO, METE CELIK**

Spatial co-location patterns represent the subsets of features whose instances are frequently located together in geographic space. Co-location pattern discovery presents challenges since the instances of spatial features are embedded in a continuous space and share a variety of spatial relationships. A large fraction of the computation time is devoted to identifying the instances of co-location patterns. We novel join-less approach for co-location pattern mining, which materializes spatial neighbor relationships with no loss of co-location instances and reduces the computational cost of identifying the instances. The joinlessco-location mining algorithm is efficient since it uses an instance-lookup scheme instead of an expensive spatial or instance join operation for identifying co-location instances. The experimental evaluations show the join-fewer an algorithm perform more efficiently than a current join-based algorithm and is scalable in dense spatial datasets.

**2.4.1 JOIN-LESS CO-LOCATION MINING ALGORITHM**

The join-less co-location mining algorithm has three phases. The first phase converts an input spatial dataset into a set of disjoint star neighborhoods. The second phase gathers the star instances of candidate co-locations from the star neighborhood set, and coarsely filters candidate co-locations by the prevalence value of the star instances. The third phase filters co-location instances from the star instances, and finds prevalent co-locations and generates co-location rules a join-less algorithm trace. Algorithm 1 shows the pseudo code. Convert a spatial dataset to a set of disjoint star neighborhoods given an input dataset and a neighbor relationship, first find all neighboring object pairs using geometric method such as plane sweep or a spatial query method using quaternary tree or r-tree. The star neighborhoods are generated by grouping the neighboring objects per each object. Shows the star neighborhoods sorted by the feature type of the center objects. Generate candidate collocations (step 4): size candidate co-locations are generated from prevalent size co-locations. Here, we have a feature level filtering of co-locations. If any subset of a candidate co-location is not prevalent, the candidate co-location is pruned.

a join-less co-location mining algorithm with neighborhood materialization. the algorithm is efficient since it does not require expensive spatial joins or instance joins for identifying co-location instances. the experimental evaluation shows the join-less method outperforms the join-based method and is scalable to dense datasets. as future work, we plan to explore methods to answer temporal questions such as how a co-location changes over time as well as methods to identify objects showing similar moving patterns.

**2.5 MINING CO-LOCATION PATTERNS WITH RARE EVENTSFROM SPATIAL DATA SETS**

**AUTHORS:YAN HUANG AND JIAN PEI, HUI XIONG**

A co-location pattern is a group of spatial features/events that are frequently co-located in the same region. For example, human cases occur in regions with poor and the presence of birds. For colocation pattern mining, previous studies often emphasize the equal participation of every spatial feature. As a result, interesting patterns involving events with substantially different frequency cannot be captured. We address the problem of mining co-location patterns with rare spatial features. Specifically, a new measure called the maximal participation ratio (maxpr) and show that a co-location pattern with a relatively high maxpr value corresponds to a colocation pattern containing rare spatial events. Furthermore, we identify a weak monotonicity property of the maxpr measure. This property can help to develop an efficient algorithm to mine patterns with high maxpr values.

**2.5.1 Maximal Participation Ratio**

there is one important observation about co-location patterns with rare spatial features, “even though the participation index of the whole pattern could be low, there must be some spatial feature(s) with high participation ratio(s) ,in pattern p the participation index is low, since west nile disease are rare compared to poor control. However, the participation ratio of “west nile” in the pattern is high.

Mining co-location patterns with rare spatial features. We first introduced a new measure called the maximal participation ratio (maxpr) and showed how this measure can be used to capture co-location patterns in spatial data sets with rare features. In addition, an algorithm called prune ax was developed to exploit the weak monotonicity property of the maxprmeasure and efficiently identify co-location patterns with rare features. Finally, our experimental results showed that the performance of the prune ax algorithm is much better than an alternative, the min–max algorithm, which is a simple extension of the prior-like solution.

**2.5.2 The Scalability of Maxprune**

The scalability of the max prune algorithm with respect to maxpr thresholds. The runtime of the max prune algorithm on data set i1m.c15.r50. As can be seen, the runtime of the max prune algorithm almost linearly increases with the decrease of maxpr thresholds. Another observation is that more features on average are in a pattern, the long the runtime we will have.

**2.6 TOPOLOGICAL RELATIONSHIPS BETWEEN COMPLEXESSPATIAL OBJECTS**

**AUTHORS:MARKUS SCHNEIDER**

Topological relationships like overlap, inside, disjoint, or meet describe purely qualitative properties that characterize the relative positions of spatial objects toward each other and that are preserved under certain continuous transformations including all affine transformations. They enable the user, for example, to ask for land parcels that are adjacent to a hazardous waste site, to explore the overlapping of a river floodplain with a proposed highway network, or to determine for a collection of cities and states which city is located within which state.

The unstructured definition purely determines the point set of a line or region. The structured definition gives a unique representation and emphasizes the component view of a spatial object. a complex point may include several points, a complex line may be a spatially embedded network possibly consisting of several components, and a complex region may be a multipart region possibly with holes. the formal framework purely employs point set theory and point set topology since us disregard shape aspects and metric properties and focus on the study of topological relationships.

**2.7 TOPOLOGICAL RELATIONSHIPS BETWEEN TWO COMPLEXES SPATIAL**

**2.7.1 OBJECTS OF DISTINCT TYPE**

Next we analyze the topological relationships between two nonempty complex spatial objects that have different types and thus different properties and dimensions. This leads to the six type combinations point/line point/region line/region line/point, region/point, and region/line. The latter three type combinations are symmetric counterparts’ of the first three type combinations so that they do not have to be treated separately. It is obvious that a number of topological relationships cannot exist. For example, higher-dimensional objects can never be located inside lower dimensional objects.

**2.7.2 CLUSTERING OF TOPOLOGICAL PREDICATES**

Based on the 9-intersection model and the predicate derivation mechanism described in we have systematically identified the topological relationships between any two spatial objects of the data types point, line, and region defined. objects of these data types have a much more complex internal structure than simple objects we have seen that topological predicates operating on complex spatial objects comprise, generalize, and extend the predicates found so far for simple object structures table ii summarizes the number of predicates obtained for each type combination.

**2.7.3 Implicit Definition Of Topological Cluster Predicates Through Clustering Rules**

Clustering rules are relaxed constraint rules which do not take into account all nine intersections of the 9-intersection matrix. Only a predicate designer but not a user can usually specify them. They must be defined in a way so that they are mutually exclusive and cover all basic topological predicates identified in for the respective type combination. as an example, we will now give a specification of eight constraint rules that is generic in the sense that it is valid for all type combinations considered.

In order to prove the mutual exclusion of the obtained topological cluster predicates as well as the complete coverage of all basic topological predicates by them, we do not argue on the basis of our definitions of clustering rules. Instead, the strategy is to match each clustering rule against all basic topological predicates, which are then associated with possible cluster predicates.

**2.7.4 Explicit Definition Of Topological Cluster Predicates Through Disjunctions**

The lists of matrix numbers in express that each clustered predicate is equal to a disjunction of basic predicates. hence, from a user perspective, a cluster predicate summarizes basic topological predicates under a common name. in a database context, a user should be able to use disjunctions of basic topological predicates for explicitly constructing cluster predicates.

Point of view, spatial applications require by far more complex geometric structures than the usual simple points, lines, and regions that can be currently found in spatial database systems, spatial query languages, and gis. In the meantime, some gis and database vendors have recognized this shortcoming and begun to incorporate more complex spatial data types into their systems. a first contribution of this article is that we have defined very general and versatile spatial data types for complex points, complex lines, and complex regions in the two-dimensional euclidean space onthe basis of point set theory and point set topology. Complex points may be composed of a finite set of isolated points, complex lines may represent spatially embedded graphs possibly consisting of several connected components, and complex regions may consist of several components where each component possibly contains holes.

**2.8 FAST ALGORITHMS FOR MINING ASSOCIATION RULES**

**AUTHORS:RAKESH AGRAWAL RAMAKRISHNAN SRIKAN**

We consider the problem of discovering association rules between items in a large database of sales transactions. we present two new algorithms for solving this problem that are fundamentally different from the known algorithms. Empirical evaluation shows that these algorithms outperform the known algorithms by factors ranging from three for small problems to more than an order of magnitude for large problems. We also show how the best features of the two proposed algorithms can be combined into a hybrid algorithm, called apriorihybrid. Scale-up experiments show that a priori hybrid scales linearly with the number of transactions.

a priori hybrid also has excellent scale-up properties with respect to the transaction size and the number of items in the database.

**2.8.1 Variation:**

counting candidates of multiple sizes in one pass rather than counting only candidates of size k in the th pass, we can also count the candidates c0k+1, where c0k+1 is generated from ck, etc. note that c0k+1 - ck+1 since ck+1 is generated from lk. This variation can pay off in these than the cost of scanning the database.

**2.8.2 Subset Function**

Candidate item sets ck are stored in a hash-tree. Anode of the hash-tree either contains a list of item sets (a leaf node) or a hash table (an interior node). in an interior node, each bucket of the hash table points to another node. The root of the hash-tree is denied tobe at depth 1. An interior node at depth d points to nodes at depth d+1. Item sets are stored in the leaves. When we add an item set c, we start from the root and go down the tree until we reach a leaf. At an interior node at depth d, we decide which branch to follow by applying a hash function to the item of theitemset. All nodes are initially created as leaf nodes. When the number of item sets in a leaf node exceeds specified threshold, the leaf node is converted to an interior node. Starting from the root node, the subset function ends all the candidates contained in a transaction as follows. if we are at a leaf, we end which of the item sets in the leaf are contained in t and add references to them to the answer set. If we are at an interior node and we have reached it by hashing the item i, we hash on each item that comes after i in tend recursively apply this procedure to the node in the corresponding bucket. for the root node, we hash on every item in t.to see why the subset function returns the desired set of references, consider what happens at the root node. for any item set c contained in transaction t, thirst item of must be in t. at the root, by hashing on every item in t, we ensure that we only ignore item sets that start with an item not in t. similar arguments apply at lower depths. The only additional factor is that, since the items in any item set are ordered, if we reach the current node by hashing the item i, we only need to consider the items in t that occur after.

**2.8.3 The AIS Algorithm**

Candidate item sets are generated and counted on-they as the database is scanned. After reading a transaction, it is determined which of the itemsetsthat were found to be large in the previous pass are contained in this transaction. New candidate item sets are generated by extending these large item sets with other items in the transaction. a large item set is extended with only those items that are large and occur later in the lexicographic ordering of items than any of the items. the candidates generated from a transaction are added to the set of candidate item sets maintained for the pass, or the counts of the corresponding entries are increased if they we recreated by an earlier transaction for further details of the ais algorithm.

**2.8.4 The SETM Algorithm**

The sets algorithm was motivated by the desire to use sql to compute large item sets. Like ais, the setm algorithm also generates candidates on-they based on transactions read from the databases tm remembers the tids of the generating transactions with the candidate item set. to avoid needing a subset operation, it uses this information to determine the large item sets contained in the transaction read. lk \_ ck and is obtained by deleting those candidates that do not have minimum support. Assuming that the database is sorted in tid order, setm can easily \_ND the large item sets contained in a transaction in the next pass by sorting Lk on tid. Intact, it needs to visit every member of lk only once in the tid order.

**2.8.5 Generation of Synthetic Data**

we generated synthetic transactions to evaluate the performance of the algorithms over a large range of data characteristic. these transactions mimic the transactions in the retailing environment. our model of the \real" world is that people tend to buy sets of items together. each such set is potentially a maximal large item set. an example of such a set might be sheets, pillow case, comforter, and rules. However, some people may buy only some of the items from such a set. For instance, some people might buy only sheets and pillow case, and some only sheets. a transaction may contain more than one large item set. for example, a customer might place an order for a dress and jacket when ordering sheets and pillow cases, where the dress and jacket together form another large item set.

We presented two new algorithms, a priori and a priory tid, for discovering all significant association rules between items in a large database of transactions. we compared these algorithms to the previously known algorithms, the ais and setm algorithms. we presented experimental results, showing that the proposed algorithms always outperform as and setm. the performance gap increased with the problem size, and ranged from a factor of three for small problems to more than an order of magnitude for large problems. we showed how the best features of the two pro-posed algorithms can be combined into a hybrid algorithm, called a priori hybrid, which then becomes the algorithm of choice for this problem. Scale-up experiments showed that a priori hybrid scales linearly with the number of transactions. in addition, the execution time decreases a little as the number of items in the database increases. as the average transaction size increases (while keeping the database size constant), the execution time increases only gradually. These experiments demonstrate the feasibility of using a priori hybrid in real applications involving very large databases.

Multiple taxonomies (is-a hierarchies) over items are often available. An example of such a hierarchy is that a dish washer is a kitchen appliance is a heavy electric appliance, etc. we would like to be able to and association rules that use such hierarchies. we did not consider the quantities of the items bought in a transaction, which are useful for some applications. Finding such rules needs further work.

**2.9 STATISTICAL ANALYSIS OF SPATIAL POINT PATTERNS**

**AUTHORS:VORGELEGT VONSTEFANIE MARTINA ECKEL**

In the following we introduce d–dimensional random point processes and their basic concepts. More information on the topics discussed in this section can be found in the existing literature. Random point processes are discussed, whereas a profound discussion of stochastic geometry including models for more general settings can be found in and finally, we mention which, without neglecting the work and contributions of many mathematicians in former times, can be regarded as the starting point of stochastic geometry as a mathematical discipline of its own right.

**2.9.1 Periodicity And Mixing**

From the point of view of application periodicity means that all the information about the point process is contained in one sample, i.e. it makes no difference if we replace averages over several realizations by spatial averages over one single realization in a larger sampling window. for theoretical aspects periodicity is necessary, e.g., for consistency properties of statistical estimators.

**2.9.2 Poisson Point Processes**

The Poisson point process is a reference model for complete spatial randomness (csr).moreover, it is a basis for the construction of more sophisticated geometric models, as well as several important point process characteristics are given by analytical formulae.

Again, Poisson point processes can be defined using the approach via the counting measures or the measurable indexing of their atoms.

**2.9.3 Characteristics for Random Point Processes**

Through point process characteristics we obtain knowledge about structural properties of point processes. a first, basic characteristic is the intensity measure which we already. The following point process characteristics offer the possibility not only to obtain qualitative knowledge about the spatial structure of appoints process, but to quantify it for specific regions of point pair distances. since they consider point pairs they are called characteristics of second order. in the following we assume that the point process s is stationary and isotropic.

**2.10 BIVARIATE NEAREST NEIGHBOR DISTANCE DISTRIBUTION FUNCTION**

the vicariate nearest neighbor distance distribution function dij(r) is the distribution function of the distance from a randomly chosen point of {(sn, ln)} with mark ito its nearest neighbor with mark j. therefore, dij(r) can be regarded as the probability that a randomly chosen point with mark ihas a distance less than or equal to rto its nearest neighbor with mark j.

**2.10.1 Minimum contrast method**

The minimum contrast method can be used for (unmarked) point patterns if an analytical formula for one of the point process characteristics of the model to be fitted is known. Actually, this method could also be used if the characteristic can be obtained by simulation, but note that such simulations unfortunately make the method time consuming and inefficient.

**2.10.2 Correction of edge effects:**

if the sampling window w is rectangular we can apply periodic boundary conditions by identifying opposite sides of w, so that a point that is located near the right edge has neighbors near the left edge, i.e. w is wrapped onto a torus. this method typically reduces bias, but increases variance.

**2.10.3 Replicated point patterns**

under the assumption those k realizations of the point process s are independently sampled, i.e. we have so–called replicated point patterns; we can use a method by to fit common parameter values to all replications. the assumption that unknown parameters are identical for all replications can be relaxed. one way to do this is to condition on the number of points in each replicated point pattern .another approach based on generalized linear mixed models is given in where the use of models with random effects allows relaxing the assumption on a common interaction parameter.

**2.10.4 Simulation of Poisson point processes**

Several stochastic point process models are based on the Poisson point process, e.g. marten cluster point processes and marten hard core point processes. As a basis to generate realizations from those, a simulation algorithm for the Poisson point process is necessary.

**2.10.5 Thinning**

For the simulation of an inhomogeneous poisson point process with a bounded intensity function λ(x) in a bounded sampling window w b0 (ird) a thinning can be applied.

**2.10.6 Birth–death case**

in the birth–death case the number of points in the sampling window w b0(ird) is determined by the activity parameter β of the gibbs point process the principle of the simulation of a gibbs point process by birth and death proposals.

**2.10.7 Bootstrapping**

Bootstrapping consists basically in an independent random re sampling of the sample data with replacement and is largely free of statistical model assumptions. it can be used to construct statistical tests and confidence intervals. The advantage of bootstrapping over analytical methods is its great simplicity.

**2.10.8 Power analysis of tests on csr**

It is important to know how sensitive the different tests react to deviations from csr, i.e. to know the power of a test. The power of a statistical test is the probability that the test rejects a false null–hypothesis, i.e. the power is the opposite of a type 2error. As power increases, the chances of a type 2 error decrease, and vice versa. The probability of a type 2 error is referred to as β.

**2.10.9 Acceptance intervals for model verification**

For the construction of point wise acceptance intervals for a distance–dependent characteristic of a given point process model, we need k realizations of this point process model. for each realization we estimate the characteristic at distances ri, i= 1, w. then, we arrange the k values for each riin ascending order.

**2.11 Spatial correlations of the change of the relative purchasing power in Baden-Württemberg**

The so–called purchasing power in a township is one of the key characteristics considered by businesses in determining site selection. it is defined here as the sum of the wages, profits and income from assets and transfers minus the direct taxes, the social security contributions, transfers, expenditures for habitation and net savings.

**2.11.1 Mark correlation function:**

A first observation is that the values for bκ(r) are above 0 for all r <50 and for all regarded time intervals, suggesting a positive correlation in the difference in relative purchasing power between townships separated by a distance r. also, w.r.t. different time intervals, the values of κ(r) are decreasing over time. this can be understood as a weakening of the positive correlations as years pass. In order to investigate these two effects in more detail, we estimate the Simpson indices.

In the application to economic data a descriptive statistical analysis of 2dimensionalmarked point patterns has been performed. we estimated the mark correlation function and the Simpson indices to describe the spatial correlations of the change of the relative purchasing power in Baden–Wurttemberg for three time intervals. the significance ofthese positive correlations for small point pair distances and its vanishing for later time intervals was also pointed out by a test on independent labeling. in the second project we detected significant positive spatial correlations of monthly stock returns in the s&p500 index. Moreover, a part of this project consisted of the comparison of two different methods for analyzing spatial correlations. The third project showed the validation of a stochastic storm track model by considering spatial correlations in the generated wind field. The tropical cyclones in the north Atlantic provide large risks for reinsurance companies. With an appropriate stochastic simulation model the hazards can be assessed more reliably. This application showed that the mark correlation function is also a useful tool for model validation.

**CHAPTER-4**

**4.3 SOFTWARE DESCRIPTION**

**About software**

**4.3.1 What is .net?**

.net is strategy of software that provides services to people any time, any place, on any device. it's an xml web services platform which allows us to build rich .net applications, which allows users to interact with the internet using wide range of smart devices (tablet devices, pocket pc's, web phones etc.), which allows to build and integrate web services and which comes with many rich set of tools like visual studio to fully develop and build applications.

**4.3.2 .net framework**

.net is a "software platform". it is a language-neutral environment for developing rich .net experiences and building applications that can easily and securely operate within it. when developed applications are deployed, those applications will target .net and will execute wherever .net is implemented instead of targeting a particular hardware/os combination. the components that make up the .net platform are collectively called the .net framework.

The .net framework is a managed, type-safe environment for developing and executing applications. The .net framework manages all aspects of program execution, like, allocation of memory for the storage of data and instructions, granting and denying permissions to the application, managing execution of the application and reallocation of memory for resources that are not needed.

The .net framework is designed for cross-language compatibility. Cross-language compatibility means, an application written in visual basic .net may reference a DLL file written in c# (c-sharp). A visual basic .net class might be derived from a C# class or vice versa. The .net framework consists of two main components:

arrowCommon language runtime (CLR)  
arrowclass libraries

**4.4 Common language runtime (clr)**

The clr is described as the "execution engine" of .net. it provides the environment within which the programs run. It’s this clr that manages the execution of programs and provides core services, such as code compilation, memory allocation, thread management, and garbage collection.

Through the common type system (cts), it enforces strict type safety, and it ensures that the code is executed in a safe environment by enforcing code access security. the software version of .net is actually the clr version.

**4.4.1 Working of the clr**

When the .net program is compiled, the output of the compiler is not an executable file but a file that contains a special type of code called the Microsoft intermediate language (msil), which is a low-level set of instructions understood by the common language run time.

This msil defines a set of portable instructions that are independent of any specific cpu. It’s the job of the clr to translate this intermediate code into a executable code when the program is executed making the program to run in any environment for which the clr is implemented. and that's how the .net framework achieves portability.

This msil is turned into executable code using a jit (just in time) complier. When .net programs are executed, the clr activates the jit complier. Thejit complier converts msil into native code on a demand basis as each part of the program is needed.

Thus the program executes as a native code even though it is compiled into msil making the program to run as fast as it would if it is compiled to native code but achieves the portability benefits of msil.

**4.4.2 Features**

Visual basic .net provides the easiest, most productive language and tool for rapidly building windows and web applications. Visual basic .net comes with enhanced visual designers, increased [application performance](http://www.startvbdotnet.com/dotnet/vbnet.aspx), and a powerful [integrated development environment](http://www.startvbdotnet.com/dotnet/vbnet.aspx) (ide). it also supports creation of applications for wireless, internet-enabled [hand-held devices](http://www.startvbdotnet.com/dotnet/vbnet.aspx).

**4.5 Powerful windows-based applications**

Visual basic .net comes with features such as a powerful new [formsdesigner](http://www.startvbdotnet.com/dotnet/vbnet.aspx), an in-place menu editor, and automatic control anchoring and docking. Visual basic .net delivers new productivity features for building more robust applications easily and quickly. With an improved integrated development environment (ide) and a significantly reduced startup time, visual basic .net offers fast, automatic formatting of code as we type, improved intellisense, an enhanced object browser and xml designer.

**4.6 Building web-based applications**

With[visualbasic](http://www.startvbdotnet.com/dotnet/vbnet.aspx) .net we can create web applications using the shared web forms designer and the familiar "drag and drop" feature. We can double-click and write code to respond to events. Visual basic .net 2005 comes with an enhanced html editor for working with complex web pages. We can also use intellisense technology and tag completion, or choose the wysiwyg editor for visual authoring of interactive web applications.

**4.7 Simplified deployment**

With visual basic .net we can build applications more rapidly and deploy and maintain them with efficiency. Visual basic .net 2005 makes "dll hell" a thing of the past. Side-by-side versioning enables multiple versions of the same component to live safely on the same machine so that applications can use a specific version of a component. Xcopy-deployment and web auto-download of windows-based applications combine the simplicity of web page deployment and maintenance with the power of rich, responsive windows-based applications.

**4.8 Powerful, flexible, simplified data access**

we can tackle any data access scenario easily with ado.net and ado data access. the flexibility of ado.net enables data binding to any database, as well as classes, collections, and arrays, and provides true xml representation of data. seamless access to ado enables simple data access for connected data binding scenarios. using ado.net, visual basic .net can gain high-speed access to ms sql server, oracle, db2, microsoft access, and more.

**4.9 Improved coding**

We can code faster and more effectively. a multitude of enhancements to the code editor, including enhanced intellisense, smart listing of code for greater readability and a background compiler for real-time notification of syntax errors transforms into a rapid application development (rad) coding machine.

**4.10 Direct access to the platform**

Visual basic developers can have full access to the capabilities available in .net framework. Developers can easily program system services including the event log, performance counters and file system. the new windows service project template enables to build real Microsoft windows nt services.

**4.11 Full object-oriented constructs**

We can create reusable, enterprise-class code using full object-oriented constructs. Language features include full implementation inheritance, encapsulation, and polymorphism. Structured exception handling provides a global error handler and eliminates spaghetti code.

**4.12 Xml web services**

Xml web services enable to call components running on any platform using open internet protocols. Working with xml web services is easier where enhancements simplify the discovery and consumption of xml web services that are located within any firewall. Xml web services can be built as easily as we would build any class in visual basic 6.0. The xml web service project template builds all underlying web service infrastructure.

**4.13 Mobile applications**

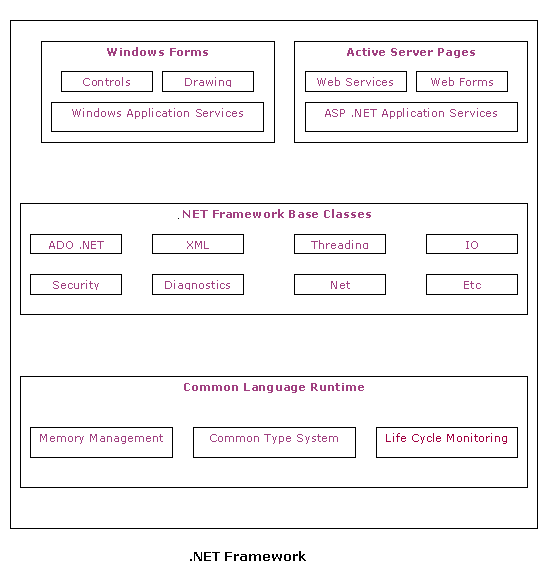
Visual basic .net 2005 offer integrated support for developing mobile web applications for more than 200 internet-enabled mobile devices. These new features give developers a single, mobile web interface and programming model to support a broad range of web devices for wap enabled cellular phones, compact html (chtml) for i-mode phones, and html for pocket pc, handheld devices, and pagers.

**4.14 Com interoperability**

We can maintain the existing code without the need to recode. Com interoperability enables to leverage the existing code assets and offers seamless bi-directional communication between visual basic 6.0 and visual basic .net applications.

**4.15 Class libraries**

Class library is the second major entity of the .net framework which is designed to integrate with the common language runtime. This library gives the program access to runtime environment. The class library consists of lots of prewritten code that all the applications created in vb .net and visual studio .net will use. The code for all the elements like forms, controls and the rest in vb .net applications actually comes from the class library.



**4.16 Common language specification (cls)**

The cls describes a set of features that different languages have in common. The cls defines the minimum standards that .net language compilers must conform to, and ensures that any source code compiled by a .net compiler can interoperate with the .net framework. some reasons for building applications using the .net framework:

* improved reliability
* increased performance
* developer productivity
* powerful security
* integration with existing systems
* ease of deployment
* mobility support
* xml web service support
* support for over 20 programming languages

flexible data access

**4.17 Microsoft access**

**4.17.1 Window-based applications**

Access is windows based application and therefore has an interface similar to Windows NT. You can cut copy and paste data from any windows application to and from access.

Since both windows and access have been developed by Microsoft corp, the two products work well together. you need to have either windows nt or windows 95 on your machine before you can install access.

**4.17.2 Large data management capacity**

Access maintains a single disk for a database and all its associated objects. The maximum size of this file can be 1 GB; access allows a maximum of 32,768 tables and objects in a database. However, access can create links to data tables in any other dbms like foxpro. Therefore, the maximum size of tables and objects is limited only by the amount of storage space that the users have.

Access can also support a maximum of 255 users at a time. These features make access a powerful dbms capable of handling large volume of data, spread across several databases and users.

**4.17.3 Importing, exporting and linking files**

Access lets you import from or export to FoxPro to FoxPro, excel, oracle and other data formats. Importing creates an access table, exporting an access table creates a file in the formal that you are exporting to. Linking means that you can use external data without creating an access table. You can link to FoxPro, excel and ASCII data.

A wizard is a utility that helps you perform complex tasks by guiding you through the process wizards ask you questions about the content, style and format of the object that you want to create, such as report or data-entry screen, and then generate the object automatically.

You need to answer the question by selecting options or making yes/no decisions. Access contains nearly 100 wizards to design database, applications, tables, forms, and reports, graphs, mailing labels, controls and properties.

**4.17.4 built in functions**

Access includes small programs known as functions that perform simple calculations or data formatting. Access includes mathematical, financial, date, time and string functions. We can use them to create expressions for calculating in the forms, reports and queries.

**4.17.5 Macros**

For these users who do not want to write programs, access provides macros. Macros let you perform common tasks without user intervention. Macros let you manipulate data, open forms and reports and automate any task that is repetitive or complex.

**4.17.6 Context – sensitive help and answer wizard**

LikeWindows NT, access provides online, context-sensitive help. You can access the help facility by invoking the help menu and selecting the topic you want help on. Access also includes a unique answer wizard that answers questions written in simple English. For example, if you want to learn how to create a database, you can invoke the answer wizard and type in your questions as “how to create a database”. Access then displays a list of topics related to database creation.

**4.17.7 Built in security**

Since access is a multi-user database, security features are built into the dbms. You can enable database security at various levels. You can assign a password for loading access so that only authorized users can use it.

you can encrypt a database to prevent unauthorized access. selective hiding at database objects such as reports, macros and forms is also possible; you can also hide selective fields in a form or report to protect sensitive data.

* windows – based application
* large data management capacity
* importing, exporting and linking external files
* wizards and builders
* built – in functions
* macros
* context-sensitive help and the answer wizard

**4.18 Microsoft windows**

Windows xp is the phenomenally successful and ubiquitous operating system software by Microsoftcorporation. it is a multitasking operating system well equipped with a graphical user interface (gui).

a task bar, always easily accessible on screen has buttons listing the currently running applications, allowing the user to easily switch between them. Windows xp is supported to make personal computers and laptops easier to learn and exploit.

**CHAPTER 5**

### FEASIBILITY STUDY

**5.1 Technology and system feasibility**

The assessment is based on an outline design of system requirements in terms of Input, Processes, Output, Fields, Programs, and Procedures. This can be quantified in terms of volumes of data, trends, frequency of updating, etc. in order to estimate whether the new system will perform adequately or not. Technological feasibility is carried out to determine whether the company has the capability, in terms of software, hardware, personnel and expertise, to handle the completion of the project.

### 5.2 Economic Feasibility

Economic analysis is the most frequently used method for evaluating the effectiveness of a new system. More commonly known as cost/benefit analysis, the procedure is to determine the benefits and savings that are expected from a candidate system and compare them with costs. If benefits outweigh costs, then the decision is made to design and implement the system. An entrepreneur must accurately weigh the cost versus benefits before taking an action.

**5.3 Cost Based Study**

It is important to identify cost and benefit factors, which can be categorized as follows:

* Development costs; and
* Operating costs.

This is an analysis of the costs to be incurred in the system and the benefits derivable out of the system.

**5.4 Time Based Study**

This is an analysis of the time required to achieve a return on investments. The benefits derived from the system. The future value of a project is also a factor. As per the cost based study this system requires the designing and implementing environment as listed below

### 5.5 Legal Feasibility

Determines whether the proposed system conflicts with legal requirements, e.g. a data processing system must comply with the local Data Protection Acts.

This system satisfies all the legal requirements and it also complying with the local data protection act.

### 5.6 Operational Feasibility

Is a measure of how well a proposed system solves the problems, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development.

This system operates well in the running environment and run as per the definition provided in the system definition.

### 5.7 Schedule Feasibility

A project will fail if it takes too long to be completed before it is useful. Typically this means estimating how long the system will take to develop, and if it can be completed in a given time period using some methods like payback period. Schedule feasibility is a measure of how reasonable the project timetable is. Given our technical expertise, are the

**CHAPTER-6**

**MODULES DISCRIPTION**

**6.1 DATABASE**

Data base is collection data in process so all details collect in the data base Automatic creation of a knowledge base content is still a research topic of interest. Information provenience de la medicine intelligent de energy eléctrica require UN analysis en un corto tiempo para una posterior toma de decisions requeridas para oftener la respect a la demand de la energy en cada empresa de distribution electrica.

**6.2 OPINION IDENTIFICATION**

Opinion identification used to identify the online data of a process thus identification for all users verify the data or information processing so easy to access for the process There are two potential sources of errors in the process of information enrichment. The first is the unstructured data origin and another is lack of accuracy of the machine learning methods used for data acquisition and information extraction.

**6.3 INFORMATION ANALYSIS ALGORITHM**

This algorithm used to unstructured of data structure process of clear information analysis of a process user can be access to data the information are collection data processing to will first reads reviews which are submitted by other customers. If that review contains the positive comments then only other customers are likely to buy that product.

**6.4 OPINION CLASSIFICATION**

The classification used to mining data are classified for the process so used to positive and negative of the process Gains complementarity by modeling both common and specific aspects across different collections, and keeping all the corresponding opinions for contrastive study.

**6.5 EVALUATION**

Thus evaluation used to data base collecting a all information so positive and negative data classify and final value collection process. Find Topical Relation between Opinion words and Opinion Target from thousands of reviews submitted by users.

**CHAPTER 7**

**SYSTEM TESTING**

**7.1Testing:-**

System testing begins by testing program modules separately followed by testing bundled modules as a unit. The total system as a single unit is tested for recovery and feedback after various major failures to ensure that no data are lost during the emergency.

**7.2 Online response:**

Online systems must have a response time that will not cause a hardship to the user. one way to list this is to input transactions on as many crt screens as would normally be used in peak hours of time the response to each online function is performed quickly to establish a true performance level.

**7.3 Recovery and security:-**

A forced system failure is indeed to test a backup recovery procedure for fill integrity. Inaccurate data are entered to see how the system responds inters of error detection and protection.

System testing is the stage of implementation, which is aimed at ensuring that the system works accurately and efficiently as expected before live operation commences. it verifies that the whole set of programs hang together. System testing requires a test plan that consists of several keys. the implementation of newly designed package is important in adopting a successful new system.

Testing is an important stage in software development. the system test in implementation should be confirmed that the system works as expected. It accounts the largest percentage of technical effort in the software development process. Testing phase in the development cycle validates the code against the functional specification. Testing is vital to the opals. the objective of testing is to discover errors. to fulfill this objective a series of test step unit, integration, validation and output testing were planned and executed. the test steps are:

**7.4 Unit& integration testing:-**

Unit testing is testing changes made in an existing or new program. This test is carried out during programming and each module is found to be working satisfactorily.

A comprehensive integration testing is carried out using integrated test plans in the design phase of the system development as guide to ensure the behavior of functions with live data.

**7.5 Validation& output testing:**

Software validation is achieved through a series of tests that demonstrates conformity with requirements. Thus the proposed system under consideration has been tested by validation and found to be working satisfactorily.

Asking the users about the format required by them tests the outputs generated by the system under consideration. It can be considered in two ways. one on the screen and other is the printed format. The output format on the screen is found to be correct as the format was designed in the system design. Output testing was done and it did not result in any change or correction in the system.

**7.6 User acceptance testing:**

The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at time of developing and making changes whenever required.

this is done regard to the following points

* input screen design
* output screen design
* online message to guide the user
* menu driven system
* format of adhoc queries and reports

**7.7 System testing:**

This is to verify that all system elements have been properly integrated and perform allocated functions. Testing is executing a program to test the logic changes made in it and with intention of finding errors. Effective testing does not guarantee reliability. Reliability is design consideration. Tests are also conducted to find discrepancies between system and its original objective, current specifications and documents.

**7.8 Performance:**

Performance is taken as the last part of implementation. Performance is perceived as response time for user queries, report generation and process-related activities.

**7.9 Software testing**

Once source code has been generated, software must be tested to uncover (and correct) as many errors as possible before delivery to your customer. Your goal is to design a series of test cases that have a high likelihood of finding errors. To do so we have techniques provide systematic guidance for designing tests that: (1) exercise the internal logic of software components, and (2) exercise the input and output domains of the program to uncover errors in program function, behavior, and performance. Resource presented in this section address the following topic categories

Software testing is the process of confirming the functionality and correctness of software by running it. Software testing is usually performed for one of two reasons:

1. defect detection
2. reliability estimation

The problem of applying software testing to defect detection is that software can only suggest the presence of flaws, not their absence (unless the testing is exhaustive). the problem of applying software testing to reliability estimation is that the input distribution used for selecting test cases may be flawed. in both of these cases, the mechanism used to determine whether program output is correct is often impossible to develop. Obviously the benefit of the entire software testing process is highly dependent on many different pieces. if any of these parts is faulty, the entire process is compromised.

Software is now unique unlike other physical processes where inputs are received and outputs are produced. Where software differs is in the manner in which it fails. Most physical systems fail in a fixed (and reasonably small) set of ways. By contrast, software can fail in many bizarre ways. Detecting all of the different failure modes for software is generally infeasible.

**7.10 Black box testing**

Black box testing focuses on the functional requirements of the software. That is black box testing enables the software engineer to drive a set of input conditions that will fully exercise the requirements for a program**.**

Black box testing is not an alternative for white box testing techniques. rather, it is a complementary approach that is likely to uncover different class of errors.

Black box testing attempts to find errors in the following categories:

* Interface errors.
* Performances in data structures or external database access.
* performance errors
* Initialization and termination errors.
* Incorrect or missing functions.
* All the above-mentioned errors were checked in the process of black box testing and the bugs found were fixed.

**7.11 White box testing:**

White box testing is some time is called glass box testing, is a test case design that uses a control structure of the procedural design to drive the test cases. using white-box testing methods, the software engineer can drive test cases that

* guarantee that logical decisions are on the true and false sides
* exercise all logical decisions are on the true and false sides
* execute all loops at their boundaries and within their operational bounds
* exercise internal data structure to assure the validity

**CHAPTER 8**

**CONCLUSION AND FUTURE ENHANCEMENT**

**8.1 CONCLUSION**

we proposed opinion mining used process of online data collection all information easy to collecting and access the process user side it’s a data structure of meaning full information process. An opinion measure using mining rules associates and snippets retrieved from a spatial database from datasets. The co-occurrence measures were computed using datasets she helped while researching and writing my paper.An information analysis algorithm too easy to data access and identify of datasets. Without her efforts and contribution, this work can be completed successfully.

**8.2 FEATURE WORK**

Also I like to thank my college who gave us access to the library and research facilities. Was proposed to identify different lexical patterns that describe the same semantic relation. Experimental results on spatial benchmark data sets showed that the proposed method outperforms various baselines as well as previously proposed soa similarity measures, achieving a high correlation with human ratings. Moreover, the proposed method improved the frequent score in a community mining.

**REFERENCES**

[1] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, “A break in the clouds: towards a cloud definition,” ACM SIGCOMM Computer Communication Review, vol. 39, no. 1, pp. 50–55, 2008.

[2] iCloud. (2014) Apple storage service. [Online]. Available: https://www.icloud.com/

[3] Azure. (2014) Azure storage service. [Online]. Available: http://www.windowsazure.com/

[4] Amazon. (2014) Amazon simple storage service (amazon s3). [Online]. Available: http://aws.amazon.com/s3/

[5] K. Chard, K. Bubendorfer, S. Caton, and O. F. Rana, “Social cloud computing: A vision for socially motivated resource sharing, ”Services Computing, IEEE Transactions on, vol. 5, no. 4, pp. 551–563,2012.

[6] C. Wang, S. S. Chow, Q. Wang, K. Ren, and W. Lou, “Privacypreserving public auditing for secure cloud storage,” Computers, IEEE Transactions on, vol. 62, no. 2, pp. 362–375, 2013.

[7] G. Anthes, “Security in the cloud,” Communications of the ACM, vol. 53, no. 11, pp. 16–18, 2010.

[8] K. Yang and X. Jia, “An efficient and secure dynamic auditing protocol for data storage in cloud computing,” Parallel and Distributed Systems, IEEE Transactions on, vol. 24, no. 9, pp. 1717–1726, 2013.

[9] B. Wang, B. Li, and H. Li, “Public auditing for shared data with efficient user revocation in the cloud,” in INFOCOM, 2013 Proceedings IEEE. IEEE, 2013, pp. 2904–2912.

[10] S. Ruj, M. Stojmenovic, and A. Nayak, “Decentralized access control with anonymous authentication of data stored in clouds,” Parallel and Distributed Systems, IEEE Transactions on, vol. 25, no. 2,pp. 384–394, 2014.

[11] X. Huang, J. Liu, S. Tang, Y. Xiang, K. Liang, L. Xu, and J. Zhou, “Cost-effective authentic and anonymous data sharing with forward security,” Computers, IEEE Transactions on, 2014, doi:10.1109/TC.2014.2315619.

[12] C.-K. Chu, S. S. Chow, W.-G. Tzeng, J. Zhou, and R. H. Deng, “Key-aggregate cryptosystem for scalable data sharing in cloud storage,” Parallel and Distributed Systems, IEEE Transactions on, vol. 25, no. 2, pp. 468–477, 2014.

[13] A. Shamir, “Identity-based cryptosystems and signature schemes,”in Advances in cryptology. Springer, 1985, pp. 47–53.[14] D. Boneh and M. Franklin, “Identity-based encryption from the weil pairing,” SIAM Journal on Computing, vol. 32, no. 3, pp. 586–615, 2003.