

## IMPLEMENTATION OF MULTI SKILL SPATIAL CROWD SOURCING PROCESSING APPROACH

**BHAGYSHREE DESHMUKH, NAMRATA KUMBHAR, SONALI NAIKNAWARE & ARATI SHINDE**

### **ABSTRACT:**

*Data mining is the computing process of discovering pattern in large data sets. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for future use. With the rapid development of mobile devices and crowdsourcing platforms, the spatial crowdsourcing has attracted much attention from the database community. Specifically, the spatial crowdsourcing refers to sending location-based requests to workers, based on their current positions. The system consider a spatial crowdsourcing scenario, in which each worker has a set of qualified skills, whereas each spatial task such as repairing a house, decorating a room, and performing entertainment shows for a ceremony is time-constrained, under the budget constraint, and required a set of skills. For developing such system different methods and techniques are used such as greedy, g-divide-and-conquer and cost-model-based adaptive algorithms to get worker-and-task assignments, the work should be done within time and under budget. The system introduce a task assignment on multi-skill oriented spatial crowdsourcing to demonstrate the efficiency and effectiveness of our MS-SC processing approaches on both real and synthetic data sets.*

**Keywords:** Multi-skill spatial crowdsourcing, greedy algorithm, g-divide-and-conquer algorithm, cost-model-based adaptive algorithm.

### **1. INTRODUCTION**

Data mining is computing process of discovering pattern in large data set. With the popularity of GPS-equipped smart and device and wireless mobile network people can easily identified and handle.

Crowd-sourcing platform assign a number of moving workers to do spatial task nearby which required

worker to move some one specified location, under budget constraint set required set of skills.

For developing such a system different method and techniques are used such as greedy, g-divide-and-conquer and cost-model-best adaptive algorithm to get workers and task, the work should be done within time and under budget.

The system introduce a task assignment on multi-skill oriented spatial crowd-sourcing to demonstrate the efficiency and effective ell also our MS-SC process approaches on both real and synthetic data set.

### **1.1 Data Mining:**

Data mining is the computing process of discovering patterns in large datasets involving methods at the intersection of machine learning, statistics and database system. It is an essential process where intelligent methods are applied to extract data patterns. It is an interdisciplinary subfield of computer science The overall goal of data mining process is to extract information from dataset and transform it into an understandable structure for further use.

1.2 LITERATURE SURVEY

Sr.No	Author Name	Paper name	Algorithm and methods	Disadvantages
1	Huiqi Hu+ , Yudian Zheng* , Zhifeng Bao++	Crowdsourced POI Labelling: LocationAware Result Inference and Task Assignment, IEEE 2016	Optimal Task Assignment	They focus on choosing labels on objects such as images and entities which do not involve the locations of tasks or workers
2.	Kun Wang, Xin Qi, Lei Shu, Der-Jiunn Deng, and Joel J. P. C. Rodrigues	Toward Trustworthy Crowdsourcing in the Social Internet of Things, IEEE 2016	System Model of Crowd sourcing in SIoT , Social-Aware Data Transmission	Time consuming, Security issues, only focus on availability threat issues relevant.
3.	Yongjian Zhao and Qi Han	Spatial Crowdsourcing: Current State and Future Directions , IEEE 2016	Unique Challenges of sc, Taxonomy of Spatial Crowd sourcing	Workers just upload the data in their vicinity at their chosen time, may not provide a complete picture of the air quality status in the city
4	Peng Cheng , Xiang Lian , Zhao Chen , Rui Fu , Lei Chen , Jinsong Han , Jizhong Zhao	Reliable Diversity Based Spatial Crowdsourcing by Moving Workers, IEEE 2015	Greedy, sampling, Divide and Conquer	It is possible that answers provided by workers are not always correct
5.	Depeng Dang, Ying Liu, Xiaoran Zhang, Shihang Huang,	A Crowdsourcing Worker Quality Evaluation Algorithm on MapReduce for	M-1 algorithm , Multi-worker evaluation scheme , M-X Algorithm	It has high computing performance and horizontal scalability. And it is suitable for large-scale worker

		Big Data Applications, IEEE 2015		quality evaluations in a big data environment.
6.	Yanrong Kang* , Xin Miao+ , Kebin Liu+	Quality-Aware Online Task Assignment in Mobile Crowdsourcing, IEEE 2015	Quality Aware Online Task assignment , Branch and Bound algorithm for HSWA	Efficient online algorithm and analyze its performance theoretically ,more comprehensive

## 2. GOALS AND OBJECTIVES

- The system which effectively handle crowdsourcing.
- In this system many task involve in one process such as,
  - 1.Constructing a house.
  - 2.Decorating a room.
  - 3.Performing entertainment shows for a ceremony.

## 3. REVIEW OF CONFERENCE/JOURNAL PAPERS SUPPORTING PROJECT IDEA

[1] “Multi Skill Oriented Spatial Crowdsourcing” ArunS.Dalvi, Prof P.N.Kalavadekarin International Journal of Innovative Research in Computer and Communication Engineering 2017.

**Description:** Presently a day there is quick improvement in cell phone gadgets with group sourcing stages, Consideration from the database group towards spatial crowdsourcing is more. Especially, the spatial group sourcing sending solicitations to specialist for their errands utilizing their current live positions. In this proposed framework, Administrator need to share and acknowledge a spatial gathering sourcing structure and each worker have some uncommon qualified course of action of aptitudes for spatial errand like building a house, painting a divider, housetop, and performing live shows for an events which is having limited obliged i.e. time and spending design and qualified fitness set.In this framework, we will contemplate and give answer for the issue of Multi Skill Oriented Spatial Crowdsourcing (MSSC). In this it will finds an imperative advantageous answer for laborer and undertaking task approach, so we can coordinate the abilities of specialist with the client characterized assignments. By utilizing this approach specialists and additionally errand client will get more advantages which is expanded with spending requirement. Henceforth, we will demonstrate that this issue is NP-hard. So we will propose a

framework or we are giving answer for the given issue with three successful methodologies, with greedy, -divide-and-conquer algorithm and overcome and cost-model based adaptive algorithm calculations to dole out qualified talented specialist for client errand which is valuable for laborers and in addition swarms. Through this expansive tests with gathering and master dataset which joins there whole information i.e. fitness set with respected worker and gathering with their profile, so we will give the beneficial and effective solution for our given issue for that we will use honest to goodness and produced datasets.

[2] "A survey on multi skill oriented spatial crowdsourcing." DalviArun S, Prof.P.N.Kalavadekar,inIJARIE 2016

**Description:** Presently a day there is quick advancement in cell phone gadgets with swarm sourcing stages, Attention from the database group towards spatial crowdsourcing is more. Especially, the spatial group sourcing sending solicitations to laborer for their undertakings utilizing their current live positions. In this general framework, Admin need to participate and accept a spatial group sourcing framework and every laborer have some extraordinary qualified arrangement of abilities for spatial errand like building a house, painting a divider, rooftop, and performing live shows for an occasions which is having restricted compelled i.e. time and spending plan and qualified range of abilities.In this framework, we will examine and give answer for the issue of multi-expertise spatial group sourcing (MS-SC), In this it will finds an imperative valuable answer for specialist and errand task philosophy, so we can coordinate the aptitudes of laborer with the client characterized undertakings.By using this approach workers as well as task user will get more benefits which is maximized with budget constraint. Hence, we are going to prove that this problem is NP-hard. So that we will propose a system or we are providing solution to the given problem with three effective approaches, with greedy, g-divide and conquer and cost-model-based adaptive algorithms to assign qualified skilled worker for user task which is beneficial for workers as well as crowds. Through this extensive experiments with crowds and worker dataset which includes there whole information i.e. skill set with respected worker and crowd with their profile, so we are going to give the efficient and effective solution to our given problem for that we will use real as well as synthetic datasets.

[3]"Task Assignment on Multi-Skill Oriented Spatial Crowd sourcing"Prof.RupaliAdhau, MeenakshiThube, Swati Narawade,Pratima Gaikwad, in International Research Journal of Engineering and Technology (IRJET) 20117

**Description:** We propose a system which will enhance the standard of results in partner air to disentangle issues by implies that of group sourcing. This structure comprises of errand administration, laborer administration, assignment dispersion, and quality investigation. Subsequently, a critical thought inside the utilization of group sourcing is to relegate adequate undertakings to each person. Additionally, to expand the standard of the outcomes.

## 4. PROBLEM DEFINITION AND SCOPE

### PROBLEM STATEMENT:

To implement the multi-skill spatial crowdsourcing approaches such as greedy algorithm, divide and conquer algorithm and cost model based adaptive algorithm, in which tasks are assign to multi-skilled workers with time constraints complex spatial task.

### SCOPE OF STATEMENT:

This application will be helpful for users who want to complete their task within time and budget. In this system User can give the feedback about his/her work and also system. System will assign task to the skilled workers.

## 5. GOALS AND OBJECTIVES

- The system which effectively handle crowdsourcing.
- In this system many task involve in one process such as,
  1. Constructing a house.
  2. Decorating a room.
  3. Performing entertainment shows for a ceremony.

## 6. MATHEMATICAL MODEL FOR PROPOSED SYSTEM

Let S be a system which assign the task to worker.

$$S = \{ \dots \}$$

Identify Input As I

$$S = \{ I, \dots \}$$

$$\text{Let } I = \{ u, w, t \}$$

The input will be the user database, Worker database and task assigned by worker.

Identify Output As O

$$S = \{ I, O, \dots \}$$

O = User assigned task will assigned to multi-skill worker according to budget and time.

Identify the processes as P

$$S = \{ I, O, P.. \}$$

$P = \{T_a, T_c\}$

$T_a$  = task will be allocated to worker.

$T_c$  = task will be completed by worker and add status as free.

Identify the failure as F

$S = \{I, O, P, F, \dots\}$

F = Failure occurs when the system fails to assign the task to worker

Identify

Success as s

$S = \{I, O, P, F, s, \dots\}$

s = success occurs when task is successfully allocated to skilled worker

Identify Initial Condition as  $I_c$

$S = \{I, O, P, F, s, I_c\}$

$I_c$  = Registration of workers

## 7. ALGORITHM

1) Greedy Algorithm :-

Algorithm Greedy (D, n)

solution  $\leftarrow$  0

for  $i \leftarrow 1$  to n do

{  $s \leftarrow$  select(D)

if ( Feasible (solution,s)) then

solution  $\leftarrow$  Union (Solution, s); } return solution

2) Divide and conquer algorithm :-

Step I :- Divide the problem into a number of subproblems that are smaller. instances of the same problem.

Step II :- Conquer the subproblems by solving them recursively.

Step III:-Combine the solutions to the subproblems into the solution for the. original problem.

3) Cost model based adaptive approach :-

Step I :- Estimate the cost by applying Greedy algorithm( $C_g$ )

Step II :- Estimate the cost by applying Divide and conquer algorithm( $C_d$ )

Step III:-Compare cost

If  $C_g < C_d$ , it select  $C_g$  otherwise select  $C_d$ .

GREEDY, g-D&C and ADAPTIVE, with a random method, namely RANDOM, which randomly assigns workers to tasks. In particular, GREEDY selects a “best” worker-and-task assignment with the highest score increase each time, which is a local optimal approach. The g-D&C algorithm keeps dividing the problem into  $g$  subproblems on each level, until finally the number of tasks in each subproblem is 1 (which can be solved by the greedy algorithm on each one-task subproblem). Here, the parameter  $g$  can be estimated by a cost model to minimize the computing cost. The cost-model-base adaptive algorithm (ADAPTIVE) makes the trade-off between GREEDY and g-D&C, in terms of efficiency and accuracy, which adaptively decides the stopping level of the divide and-conquer.

## 8. EFFICIENCY ISSUES

There is very rear case of failure is if any error occurs first time in team, no one faces this error and couldn't resolve it, Then system will fail for this fresh error.

## 9. OUTCOME

This application will help users to get on time service in critical situation where they want also it is best option for all user to get worker faster, normally they won't get worker fast to complete task.

## 10. APPLICATIONS

- 1) Construction
- 2) Wedding planning

## 11. TECHNOLOGIES USED

### 11.1. SQL Server 2008:

Microsoft SQL Server is a Relational Database Management System (RDBMS) developed by Microsoft. It is designed to run on a central server, so that multiple users can access the same data simultaneously. Generally users access the database through an application.



11.1.1 Sql Server Components:

1. Protocol Layer: Protocol layer implements the external interface to SQLServer.TDS is an application layer protocol, that is used to transfer data between database server and a client.
2. Data Storage: The main unit of data storage is a database, which is a collection of data. The data in the SQL Server database is stored in primary data files with an extension .mdf and Secondary data files, with an extension .ndf extension are used to store optional metadata. Log files in SQL Server are recognized with the .ldf extension.
3. Buffer Management: SQL Server buffers pages in RAM to minimize disc Input/Output. A 8 KB page could be buffered in-memory and the set of all pages currently buffered is called the buffer cache. On the basis of available memory, SQL Server decides how many pages will be cached in memory. The buffer cache is managed by the Buffer Manager.
4. Logging and Transaction: SQL Server uses transaction to make sure that any operation either totally completes or is undone if fails, but never leaves the database in an intermediate state. Any changes made to a page will update the in-memory cache of the page and simultaneously all the operations performed will be written to a log, along with the transaction Ideas log entry is recognized by an increasing Log Sequence Number (LSN) which makes sure that no event overwrites another event.SQL Server makes sure that the log will be written onto the disc before the actual page is written back.
5. Concurrency and locking: When multiple users update the same data or attempt to read data that is in the process of being changed by another user. In SQL Server have two modes of concurrency control-pessimistic concurrency and optimistic concurrency. In pessimistic concurrency control, SQL Server controls concurrent access by using locks (shared or exclusive).In Optimistic concurrency control, a new version of a row is created whenever the divot is updated. Both the versions of the row are stored and maintained into a system defined database Temp db.

**12. RISK ANALYSIS**

**The risks for the Project can be analyzed within the constraints of time and quality**

Risks	Category	Probability	Impact
Computer Crash	TI	70%	1
Late Delivery	BU	30%	1
Technology Will not Meet Exception	TE	25%	1
End Users Resist System	BU	20%	1
Changes In Requirements	PS	20%	2
Lack Of Development Experience	TI	20%	2
Lack Of Database Stability	TI	40%	2
Poor Quality Documentation	BU	35%	2
Deviation From Software Engineering Standards	PI	10%	3
Poor Comments In Code	TI	20%	4

- Impact Values:
  1. Catastrophic
  2. Critical
  3. Marginal
  4. Negligible

### 13. USAGE SCENARIO

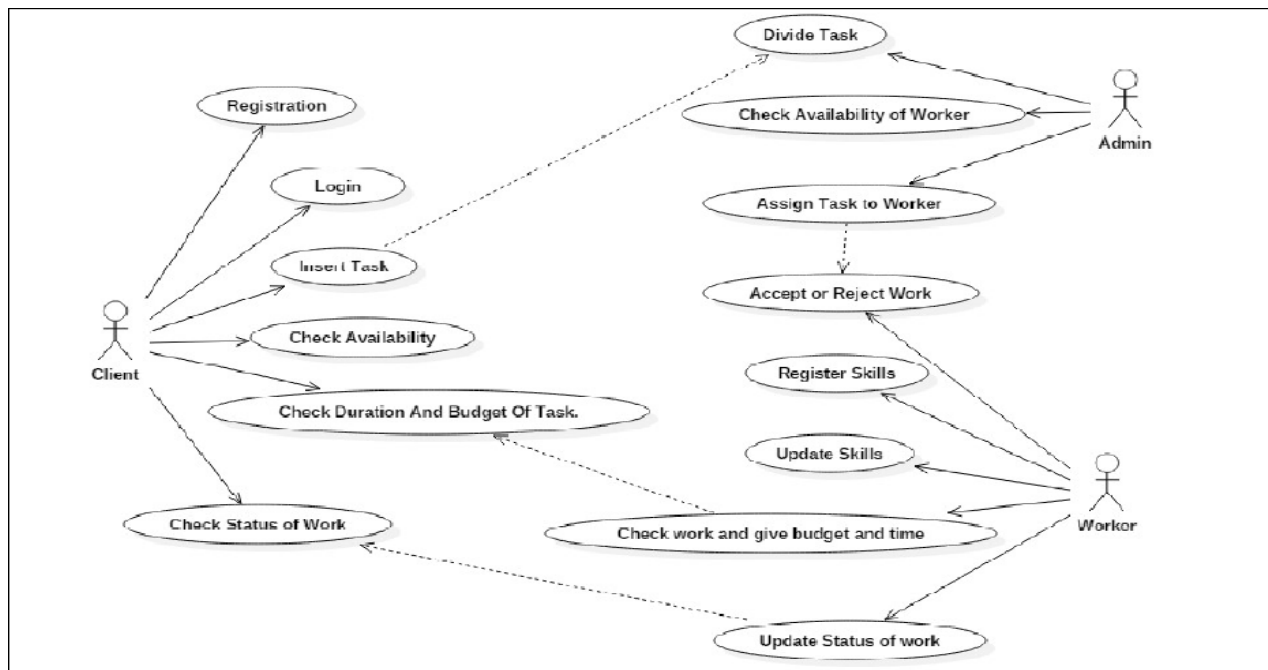
This section provides various usage scenarios for the system to be developed.

#### 13.1 User profiles

1. User
2. Server

#### 13.2 Use Case Diagram:

A use case is a methodology used in system analysis to identify, clarify and organize system requirements. The main purpose of use case diagram is to show who interacts with your system and the main goal they achieve with it.



**Fig: Use Case Diagram**

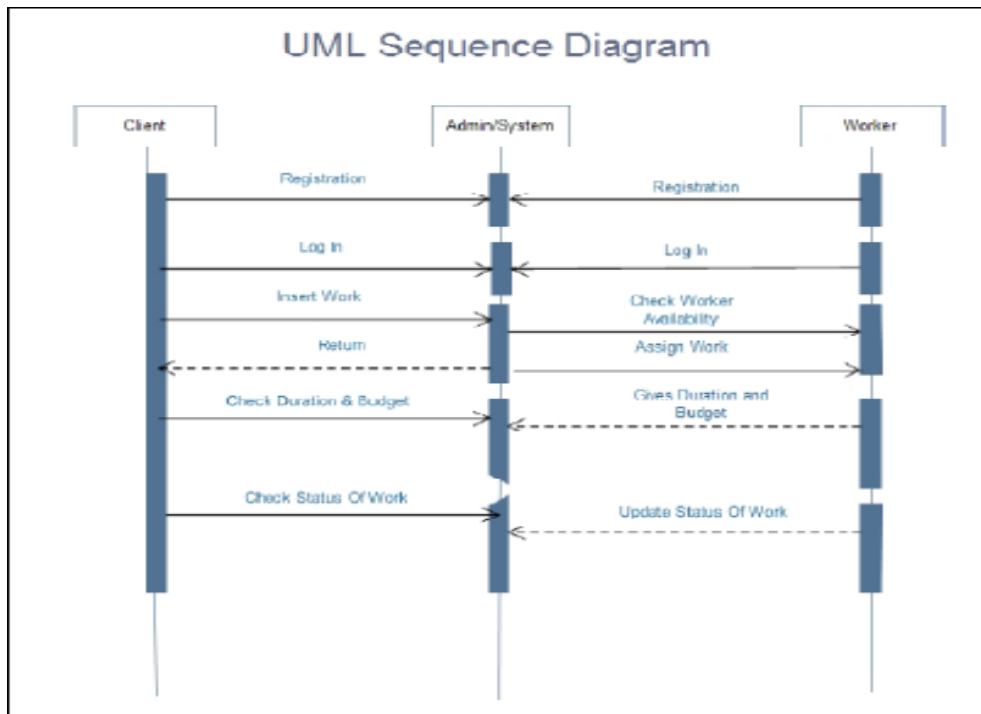
## 14. FUNCTIONAL MODEL AND DESCRIPTION

A description of each major software function, along with data flow (structured analysis) or class hierarchy (Analysis Class diagram with class description for object oriented system) is presented.

### 14.1 Sequence Diagram:

A Sequence diagram is an interaction diagram that shows how objects operate with one another and in what order. It is a construct of a message sequence chart. A sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the Logical View of the system under development. Sequence diagrams are sometimes called event diagrams or event scenarios.

Sequence Diagram is as follows –

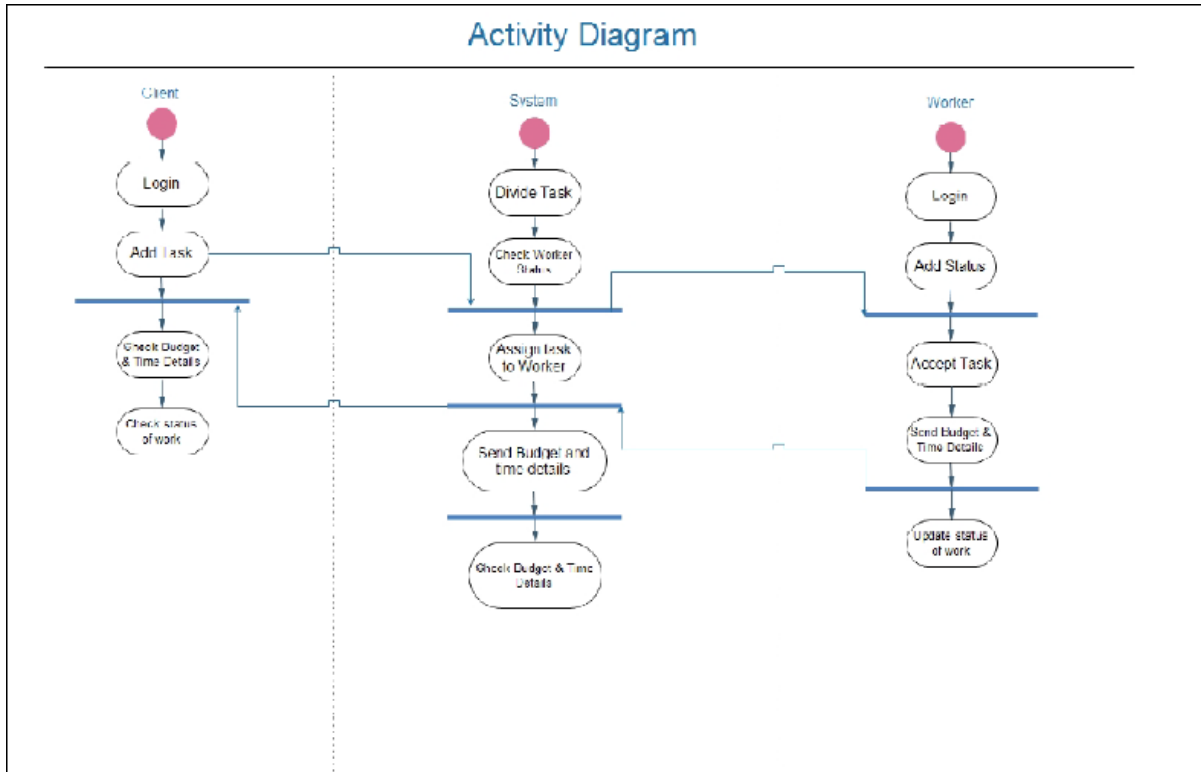


**Fig: Sequence diagram**

### 14.2 Activity Diagrams

Activity diagrams are graphical representations of work flows of stepwise activities and actions with support for choice, iteration and concurrency. An activity diagram shows the overall flow of control. Activity diagrams are

constructed from a limited number of shapes, connected with arrows. The most important shape types: Rounded rectangles represent activities; Diamonds represent decisions; Bars represent the start (split) or end (join) of concurrent activities. Figure 6.4 represents the activity diagram of the system.



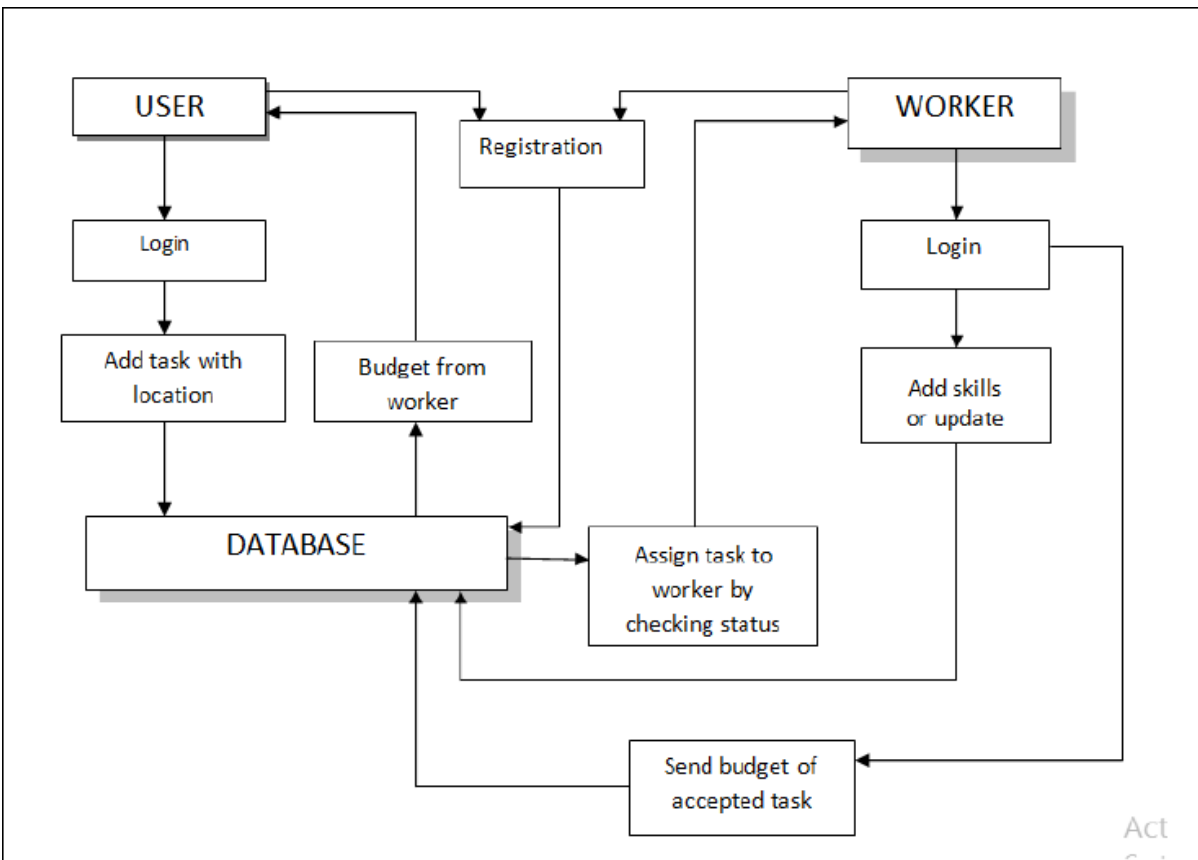
**Fig: Activity diagram**

## 15. SOFTWARE ARCHITECTURE

Software architecture refers to the high level structures of a software system, the discipline of creating such structures, and the documentation of these structures. It is the set of structures needed to reason about the software system. Each structure comprises software elements, relations among them, and properties of both elements and relations.

### 15.1 ARCHITECTURAL DESIGN

Client by using this system Clients can creates their account and take information regarding the nearest appropriate Mechanic/Technician and send him an appointment request.



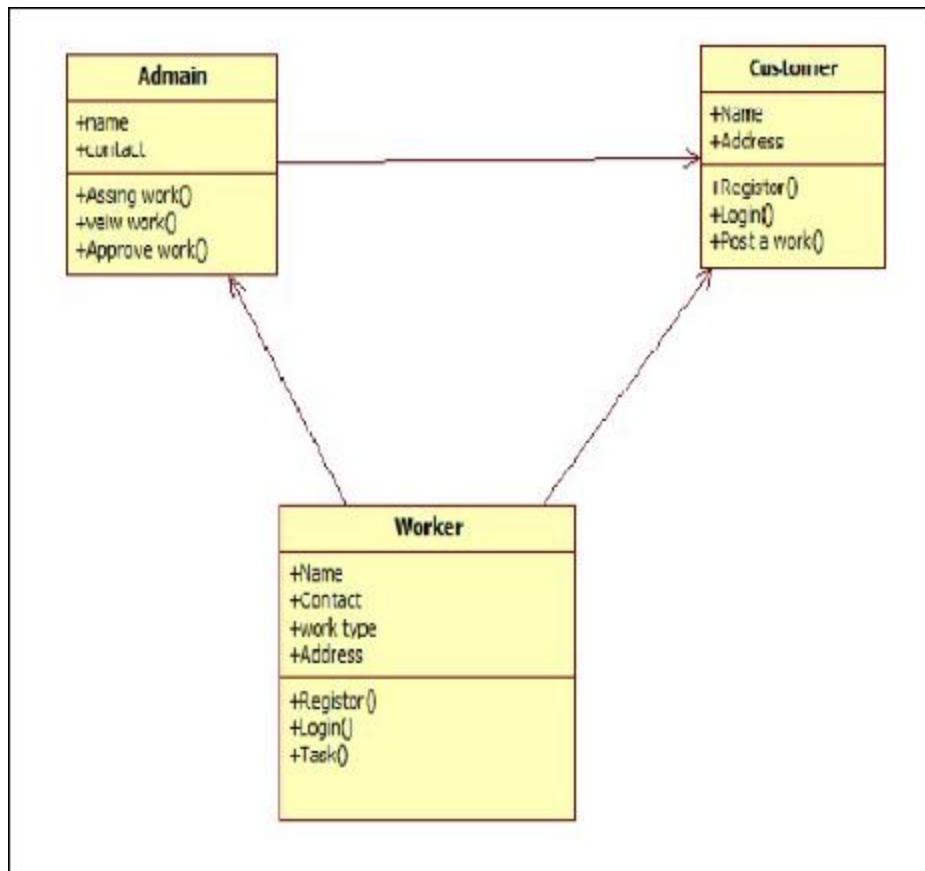
**Fig : Architecture diagram**

## 16. COMPONENT DESIGN

A component design is design specification for one of these adaptable components. Application engineers, using generation procedure may adapt and compose a set of these components to implement certain work products.

### 16.1 Class Diagram:

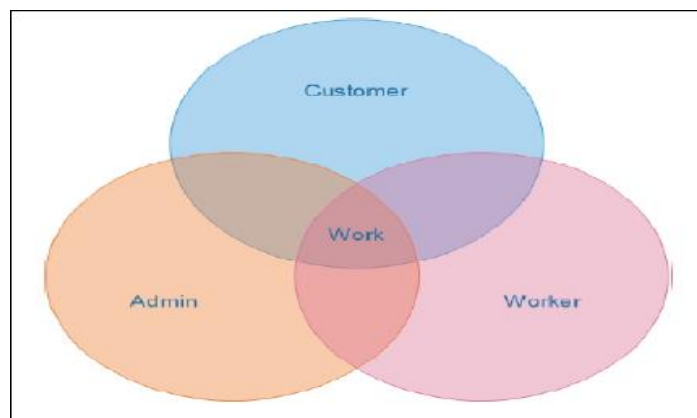
In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system’s classes, their attributes, operations (or methods) and the relationships among the classes. The class diagram is the main building block of object oriented modeling. It is used both for general conceptual modeling of the systematic of the application, and for detailed modeling translating the models into programming code.



**Fig : Class Diagram**

16.2 Venn Diagram:

A Venn diagram or set diagram is a diagram that shows all possible logical relations between a finite collection of different sets. Venn diagrams were conceived around 1880 by John Venn. They are used to teach elementary set theory, as well as illustrate simple set relationships in probability, logic, statistics, linguistics and computer science.



**Fig: Venn Diagram**

## 17. CONCLUSION

This system proposed by the problem of the multi-skill oriented spatial crowdsourcing (MS-SC), which assigns the time-constrained and multi-skill-required spatial tasks with dynamically moving workers, such that the required skills of tasks can be covered by skills of workers and the assignment score is maximized.

This system prove the processing of the MSSC problem is NP-hard, and thus this system proposed three approximation approaches (i.e., greedy, g-D&C, and cost-model-based adaptive algorithms), which can efficiently retrieve MS-SC answers. Extensive experiments have shown the efficiency and effectiveness of system proposed MS-SC approaches on both real and synthetic data sets.

## 18. REFERENCES

- [1] Zhe Yang, Zhehui Zhang, Yuting Bao, Xiaoying Gan, Xiaohua Tian, Xinbing Wang. On Tac: Online Task Assignment for Crowd sourcing, In IEEE 2016
- [2] Habibur Rahman, Senjuti Basu Roy, Saravanan Thirumuruganathan, Sihem Amer-Yahia, Gautam Das. Task Assignment Optimization in Collaborative Crowdsourcing. In 2015 IEEE.
- [3] Ioannis Boutsis, Vana Kalogeraki. On task assignment for real-time reliable crowdsourcing. IEEE 2014.
- [4] Khanh-Hung Dang and Kim-Tuyen Cao. Towards Reward-based Spatial Crowdsourcing. ICCAIS 2013.
- [5] Ioannis Boutsis, Vana Kalogeraki, Crowdsourcing under Real-Time Constraints. 2013 IEEE
- [6] Z. B. G. L. J. F. Huiqi Hu, Yudian Zheng and R. Cheng. Crowd-sourced poi labelling: Location-aware result inference and task assignment. ICDE 2016.
- [7] S. S. Kanhere. Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces. In MDM 2011.