IR using A* algorithm to Improve Performance

B. Haritha[1], MD. Jareena Begum[2]

[1]Department of Computer Science and Engineering, Bapatla Women's Engineering College, A.P
[2]Department of Computer Science and Engineering, Bapatla Women's Engineering College, A.P

ABSTRACT

Now a Days Information Retrieval plays a vital role. Information retrieval (IR) is the process of obtaining information from various resources by performing searches. This information may be a document or image or Meta data or sound etc. The IR not only deals with searching but also deals with the representation of information, storage and organization of knowledge. But the information is different from the data. The term information means organised and processed data with a specific meaning where as data means raw and unprocessed material it does not carry any specific meaning. The large volumes of data is called Big Data. It is both structured, unstructured. So it requires proper techniques and strategies to deal with that data. To retrieve that raw data it is cost effective, because of dealing with unstructured data, managing the data, controlling and storing that data. For this purpose in this paper use AI tools to retrieve the data most efficiently. It is Faster, accurate, less cost requirement compare to previous techniques.

I. INTRODUCTION

The intelligence exhibited by machines or software is known as **Artificial intelligence** (**AI**). As human being's brain perceives its environment and then takes decision based on perception. Similarly the intelligent agent in artificial intelligence can understand the environment called perception and then able to take decisions, through the artificial neural network. Artificial Intelligence research is divided into number of subfields that sometimes may fail to communicate with each other. Some sub fields are used to solve a problem with help of some AI tool. Knowledge, Reasoning, learning, planning, NLP (for communication), perception, the ability to move and manipulate the objects are the parts of **Artificial intelligence.** The main goal of Artificial Intelligence is to make intelligent systems like working of human brain. Basically the Big data is a huge collection of digital information which is growing continuously through number of channels like social media, e-mail, customer and B2B transactions, user generated content like photos and videos, GPS-equipped mobile phones, social networking apps (facebook, whatsapp, hike etc), system logs of all kinds.

Today it has become an essential part of the technology industry, providing the heavy lifting for many of the most challenging problems in computer science. Today it becomes an essential part of the technology industry and providing the heavy lifting for most challenging problems in computer science field. These sources giving a continuous stream of digital information without any limit hence it is called as "Big data". Big data act as a driving force that will affect organization's profit. Through Big data achieve a better understanding of organization's customers, employees, partners and operations. Inspite of being so valuable, all organizations were not able to afford it, because of its unstructured data, infrastructure and more cost expenses.

For implementation of Big Data all the departments in an organization have to undergo with many changes which are tedious job and also the users across the organization must have a power to work with the diverse data sets through the self-service tools to acquire a Big data productivity. After the adoption of the Big data,

the data analysis process will become a very important part of the business while it is considered as separate part earlier and not given so much importance. The Big data analysis is not only facilitates the company to acquire a better understanding, but also what else is possible for that.

The main objective of a data scientist is to convert raw data into valuable information that information yield rich, more accurate, wide ranging and actionable insight which can help address the customer needs, operational risks and performance opportunities within the organization and extended supply chain.

The main aim of this paper is to retrieve the large amounts of information with efficient manner using AI tools and techniques. There are many search techniques in Artificial Intelligence. Those are

1) Informed search technique
2) Uninformed search technique

Informed search: In the case of Informed search technique, it has additional information related to goal state which helps more efficient in searching. It compulsory gives the solution and it is smart to visit nodes.

Uninformed search: In the case of Uninformed search technique, it need not have any additional information related to goal state. It may or may not give the solution. It is blind search to visit the nodes.

Therefore compare to uninformed search technique, informed search technique is better in case of performance. There are several techniques under informed search techniques. Those are

- Best first search
  - ➢ Greedy best first search
  - ➢ A* search
- Hill climbing search
- Constraint satisfaction problem.

- Generate and test
- Means end analysis
- Problem reduction

Among these, we should use A* search for retrieval of information. A*is a best algorithm in informed search techniques or heuristic search techniques.

II. PROPOSED SYSTEM

The proposed system is consists of following steps.

1) Input the user query.
2) Apply the Artificial Intelligence based algorithm
3) Returning the data file.
4) Input the user query to the expert system
5) Analyzer: Searching the user query if it was existed in the historical file or not
6) Return the results.

There are several phases in this system .the above steps are explained in these phases.

Phase 1: This phase includes select the database file that is to be inputted to the AI algorithm. The inputted database file can be contained any type of data.

Phase 2: This phase comprises the selection of the best artificial intelligence search algorithm. The AI based search algorithm depends on the best informed search technique through the problem solving agent.

In *informed* search, a heuristic or guess is used as a guide that is for better performance to get the goal state. Instead of visiting the search tree blindly, perform evaluation function at each and every node. Based on the evaluation function visit the best node then go best next. This node path is then expanded and the process is repeated until to get the goal node. For this process prefer A* Search algorithm. It is a form of Best first search.

In order to move towards the goal, the evaluation function must estimates the cost, to reach the closest goal state from a given state. This can be done based on some knowledge about the problem domain. The best first search and A* search follows DFS procedure internally from starting node onwards.

Phase 3: In this phase, it displays the results of the AI based algorithm.

Phase 4: Here the users entering the query to the expert system.

Phase 5: This phase comprises analyzing the user query, if the user query exists or not in the historical file.

Phase 6: It is the final phase consists of displaying the answers to the query through the expert system and saving that user query in historical file and then return the query answer to the user.

Figure 1: Proposed System Architecture

Our methodology is based on the intelligent agents. The agent observes the environment through sensors and acts upon the environment using actuators and directs its activity towards achieving the goal. In this model we are using the two types of agents. i.e.,

- Model based agents
- Goal based agents.

A model-based agent can handle only a partially observable environment. Its current state information is stored inside the agent. This knowledge is about "how the world works" is called a model of the world, hence its name is "model-based agent".

Similarly the goal-based agents further expand on the capabilities of model-based agents, by using the "goal" information. Goal information describes the situations that are desirable. It allows the agent to select the best possibility to reach the goal state. Basically the main objective of using agents is to keep track of the world state as well as a set of goals are trying to achieve a goal and chooses the best action that will lead to achieving the goal.

ALGORITHM

SIMPLE-PROBLEM-SOLVING-AGENT (PERCEPTS)
returns an action.
Inputs: precepts, a percept
Static: Sequence, an action sequence, it is initially empty, some description about the current world
States:
 Goal State, a goal, initially a
 null Problem, a problem
 formulation
States← UPDATE_STATE (state, percept)
if sequence is empty **then do**
 Goal_ State ← FORMULATE –GOAL (state)
 Problem ←FORMULATE PROBLEM (states, Goal_State)
Sequence← SEARCH (problem)
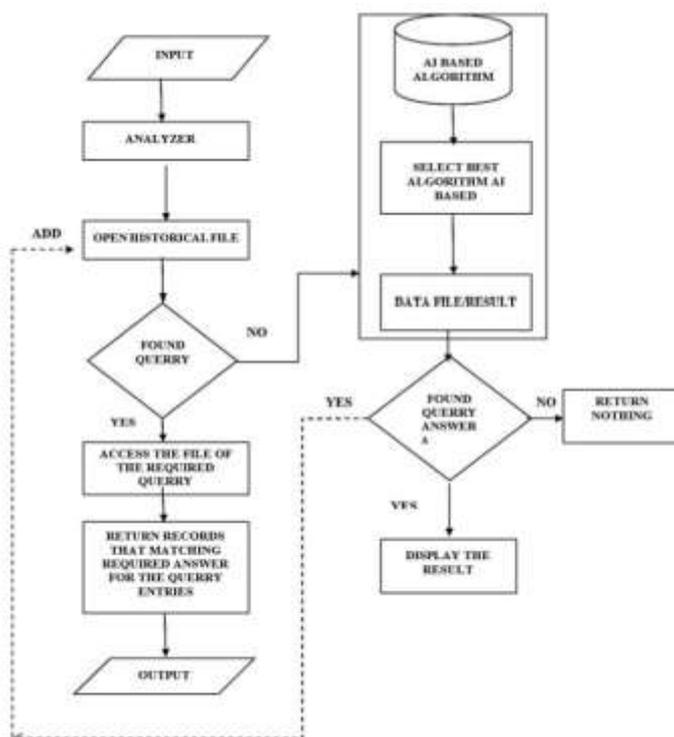Action ← FIRST (sequence)
Sequence← REST (sequence)
return action

Problem:- A problem can be defined formally by the five components:

Initial state: - It is an initial state of an agent.

Actions:- A clear description of the possible actions available to an agent. Given a particular state called's', ACTIONS (s) returns a set of actions that can be executed in state "s".

Transition model: It tells what action is done in current state and that reflects to which particular next state among numbers of nodes. It specified by a function RESULT (state, action) that returns the state that results from doing some action "a" in state current state "s". Here we will also use the term called successor which refers the next state from the current state by performing some possible action "a".

State space:- The set of all states reachable from the initial state to goal state. Initial state, actions, and transitions implicitly define the state space of that problem.

The state space forms a Graph or network which having the sates, links from one node another and actions upon the links.

Function

RECURSIVE-BFS (problem)
  return a solution, or failure RBFS (problem, MAKE-NODE (INITIAL-STATE [problem]), ∞)

Function RBFS (problem, node, flimit)
 return a solution, or failure and a new fcost limit

If GOAL –TEST [problem] (state) **then return** node Successor ← EXPAND (node, problem)

If successor is empty **then return** failure, ∞

For each s **in** successors **do**
f[s]←max(g(s) + h(s), f[node])
Repeat

best← the lowest fvalue node in successors **if** f[best] > flimit **then return** failure, f[best]
alternative ← the second lowest fvalue among successors

result, f [best] ← RBFS (problem, best, min (flimit, alternative))

if result ≠ failure **then return** result

It works by maintaining on the recursion stack. The complete path from the current node to goal node, being expanded and all the immediate siblings of current node on that path, along with the cost of a best node in the sub tree expanded below each sibling. When the cost of the current node exceeds with some other node in the previously expanded tree, then the algorithm backs up to their deepest common ancestor, and continues the search procedure down to the new path. In effect, this algorithm maintains a separate threshold value for each sub-tree deriving from the current search path.

RBFS or A* is optimal (if the heuristic is admissible) and robust, but it suffers from excessive node regeneration due to low memory profile, which effects a long processing time. Given enough time, though, it can solve problems.

III. CONCLUSION

Future use of Big Data with the help of AI is very bright. The use of artificial intelligence would lead to production of machines and computers, which are much more advanced than present systems. Researchers were continuously working to handle increasing of data as well as to convert it into valuable assets. Some-day computer will be able to do any type work. The Speech recognition systems will reach much higher levels of performance and able to communicate with humans, using text and voice, in unstructured English. Some day there will be a great future for expert system applications in all aspects like health care, in both clinical and administrative areas,

improving patient conditions and in allocation of social, financial, and other resources. However, when it comes to the question of by using the Artificial Intelligence creating machines, which are more intelligent than humans, no one seems to have the answer. In addition to that, even if it is possible, calculate the amount of time it will take cannot be predicted. It is also expected to have human brain features like learning from perception, cognition, and experience.

IV.REFERENCES

[1]   Jump up ^ "Manufacturing: Big Data Benefits and Challenges". TCS Big Data Study. Mumbai, India: Tata Consultancy Services Limited. Retrieved 2014-06-03.

[2]   Avneet Pannu; "Artificial Intelligence and its Application in Different Areas", ISSN: 2277-3754 ISO 9001:2008 Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 4 issue 10 April 2015.

[3]   Bart Selman,(moderator) AT & T, Rodney A. Brooks MIT, Thomas Dean (Brown University), Eric Horvitz (Microsoft), Tom M. Mitchell (CMU), Nils J. Nilsson (Stanford university) "Challenge problems for Artificial Intelligence" ; To appear in Proc. AAAI-96.

[4]   Avneet Pannu, M. Tech Student,
 DAV Institute of Engineering and Technology, Jalandhar India   "Artificial Intelligence and its Application in Different Areas" International Journal of Engineering and Innovative Technology (IJEIT) Volume 4, Issue 10, April 2015.

[5]   Juan Ramón Rabuñal Dopico University of A Coruña, Spain Julián Dorado de la Calle University of A Coruña, Spain Alejandro Pazos Sierra University of A Coruña, Spain" Encyclopedia of Artificial Intelligence"

[6]   Imen Akermi IRIT, Université Toulouse IIIToulouseFrance, Gabriella Pasi Università degli Studi di Milano-BicoccaMilanItaly,Karam Abdulahhad ,
GESIS - Leibniz Institute for the Social SciencesCologneGermany "Information Retrieval and Artificial Intelligence" Spinger link .

[7] Abdulahhad K (2014) Information retrieval (IR) modeling by logic and lattice. Application to conceptual IR. Theses, Université de Grenoble.

[8] Abdullah HS, Hadi MJ (2014) Artificial bee colony based approach for web information retrieval. Eng Technol J 32(5):899–909.