This chapter looks into the motivation and inspiration for this thesis. In this digital era special focus should be laid on information dissemination. The way in which knowledge is gathered, filtered and presented is drastically changing. The problems that have aroused due to information overload are explained in detail. Under unique scenarios, simple solutions to present situation will work optimally. This is because they have omniscient view of the entire ambience and can use that view to find the best solution. However, in real world conditions, computational and bandwidth power are limited, communication and data processing takes time. In these scenarios the weaknesses of present day approach quickly become apparent. Strategies that can compensate for these and more challenging conditions are looked into in this thesis.

Information overload problems exist in almost all domains such as tourism industry, healthcare systems, education portals etc. This problem is even more prominent in farming industry since target audience education and digital information consumption levels are usually low. This provides good scope for this research but at the same time provides very unique challenges. Multi-agent systems work very well in solving complex real world problems. Integration of third party tools along with technologies like Ontologies and fuzzy approach in agents provide intelligence and support to handle information overload problems. Development and deployment process can be eased by the usage of established agent platforms like JADE.

In Section 1.1 of this chapter the background of the research topic is discussed. Technologies that are used in this research are mentioned briefly in Section 1.2. Objectives covered in this research work are listed in Section 1.3. Section 1.4 presents the outline of this thesis and Section 1.6 concludes the chapter.
1.1 Background

We live in an information age. Information technology revolution has profound changes in the nature of our world. The benefits of information technology are a major driver, not just for improved quality of life, but also of economic growth [94]. Information is a major development resource along with the ranks of human, natural and financial resources [48]. The revolution has not just been limited to data on the web but communication options have literally exploded. Communication with someone located anywhere in the world has been possible with IT revolution on a real-time visualization basis [99]. Text and video based chat’s as well as the near instantaneous method of email have made communication lot easier. When letters were the main mode of communication one has to post it and wait for a response. With modern computers having built in web camera’s and high speed internet access becoming cheaper, long distance phone calls are also slowly becoming unnecessary [53].

Knowledge becomes more and more dynamic as experiences accumulate [73]. It is continuously revised with new additions while at the same time other knowledge passes into obsolescence. With the use of technology it has become much easier to create new bodies of knowledge. The ease of duplication and transmission of data have increased information overload problems many fold. The term “Information overload” was first used by futurologist Alvin Toffler [111] in 1970. Over the years the problem has been increasing exponentially. The ease with which information can now be created and passed there is no easy end to this complexity. Users who can deal with it can take advantage of all the available knowledge from different sources.

This information overload problem is not just in farming industry it exists in various domains but we are looking at this industry in particular. Most of the farmers in India don’t even have basic education and they might not even know basic computer operations. If it is up to farmers to enter the right keywords and find right information from the Web they will be lost. As now a day’s the Internet has emerged and established as a major source of information delivery, more focus should be laid on information search and retrieval methods. More emphasis should be given to the conditions in which people will be using these applications.
Information is an essential key factor to improve agriculture production. In this new technology age farmers need to get information at the right time to improve and market their yield better. In the last century the only source of information for farmers was by word of mouth. Before the advent of the Internet the main sources of information to farmers were through newspapers, TV, Radio, cooperative societies, village fairs, government demonstrations etc. With the advancements in information storage and retrieval methods, most of these sources do not exist anymore or their significance has greatly diminished.

This electronic revolution has affected farming industry also but not to the extent that it did to business world. USDA (United States Department of Agriculture) recently reported that the since 1997 the usage of computers on farms has increased from 38 to 55, while the Internet use on farms has advanced from 13 to 43 percent [107]. Farmers in some of the countries where electronic resources are more readily available are beginning to embrace e-business and successfully trade on-line. Internet offers additional benefit of attenuating some of the constraints on a farmer’s ability to receive and manage information. Regardless of where the farm is located or when the information is used internet makes data resources available around the clock. Many agricultural groups, teachers, farm organizations, researchers, and extension agents have taken an keen interest in Internet use in agriculture [107].

The Internet is not embraced on Indian farm lands as much as in the west. Farm lands in India are fragmented and have very weak infrastructure. The involvement of numerous intermediaries makes it harder to take the technology revolution further. Typically rural Indian farmers sell their crops in local mandi’s. They lack access to market information and only have an approximate idea of price ranges and have to accept whatever price is offered to them by the traders in the mandi’s. Typically poor and often illiterate, these farmers have very limited access to information and education. They are lacking access to improved farming techniques, which could enhance their yield [37].

Information is in various formats (audio, visual, audio-visual, text) and this has to reach farmers at a low cost. Farmers are apprehensive about adopting new farming practices or crops. Farmers are used to certain farming techniques as an adherence to tradition, sometimes dating back several generations. Since crop yield is their main source of income
they don’t like to risk their family’s well being. They must be shown enough proof and given proper guidance that new practices will result in better yield which means a better standard of living. Figure 1.1 illustrates some of the reasons why users have not embraced the Internet. As such, a major hurdle for the acceptance of new practices is one of education and trust. Youth are running away from farm lands. Involving them in the education process and usage of technology will create microenterprise opportunities for them [36].

As time passes by weather patterns are changing, the quality of soil, water resources and the whole medium of agriculture is changing. There are lots of factors which affect the farming yield [3]. Quality of lives of farmers greatly improves when they can grow more food and earn more income. They can also be self-sufficient and live better lives. When small farmers are able to earn better, they can take better care of their families. Children can be sent to school, will get access to better health care and can also invest more in their farms. Improvements in agricultural productivity create social and economic ripple effects. This makes their communities economically stronger and more stable [17]. That is what is covered in the paper written by S.C. Mittal [98] where he talks about the indirect benefits of IT in empowering Indian farmers. Farmers require timely and reliable sources of information for taking proper decisions. Right now farmer depends on information from conventional sources which are slow and unreliable. Environment and markets are changing rapidly and farmers need fast reliable information to remain competitive.
Searching for “increasing rice yields” in any search engine returns 18,200,000 Web results, 17,700,000 video links, 3,190 news links and 896,000 blog references. Except for a very few broad categories like web results, videos, new links, books, blogs etc the results are not organized by any subject relevance. An educated computer savvy individual with good Internet connection can open a few leads and figure out a good strategy to deal with the problem. Often it is not a straight forward process and may take up quite a bit of time. This is when we have a specific problem and trying to find out the solution for it. When the problem cannot be specified properly the chances of finding the right solution are also slim.

The technologies used for publishing and distributing information and knowledge on the Web have improved significantly. Data retrieval from the Web for simple queries works great but it is hard to find answers to questions that require more than a keyword search. Information on the Web is growing at an exponential rate and is organizing itself based on the context rather than the content. Page layout issues are given more importance. This universal knowledge base has a bigger untangled problem of interpretation of its contents. Although widely used and acknowledged for the advantages it offers the increasing popularity shows major drawbacks [58].

There are 16 million farming websites on the internet today. The problem is to find genuine ones to get information from the sea of knowledge. As most farmers are not Web or technology savvy it takes lot more effort to navigate them through the sea of information. It is a laborious and time consuming process to query multiple data sources manually on the Internet for farmers. This gets extremely challenging when resources like computers, Internet bandwidth and power are severely limited. Traditional information systems work well for simple queries but do not scale well to the diverse, vast, and the growing number of farming data sources. With the help of search engines users can search through large numbers of data sources. Search engines retrieve data but do not provide any capabilities for locating, processing, combining, and organizing information [40]. Also search engines don't take context of the pages into consideration. Not just the information overload the number of communication channels as shown in Figure 1.2 has also increased like individual web pages, blogs, forms, social networking sites...etc. There is no easy way to verify the integrity and quality of the information users are presented with. The information overload
is not just a problem for end users but also for architects who design information systems [8].

Figure 1.2: Present Sources of Information

1.2 Objectives

The main objective of this research is to handle information overload problem. A Web based portal developed using multi-agent framework provides needed support to farmers to aid in their activities. Agents are developed to combine knowledge from local databases and also can tap into the vast intelligence available on Web. The usability of the system is enhanced by agents performing certain actions preemptively on user’s behalf. The prototype was developed for farming industry but could be easily extend to other fields such as tourism industry, health industry, education portals etc using the generic framework designed as a part of this research work.

When data is everywhere around and scattered in different mediums one of the primary goals of agents is to optimize the storage and retrieval of data. The advantages and short comings of storing content in knowledge bases and retrieving them always from the Web are considered while developing the prototype. Usage of technologies like fuzzy logic helps build interfaces where human like reasoning and behavior are required. Along with using agents to optimize data retrieval methods and taking care to organize the retrieved content there is also a need to weed out unwanted data. To achieve this personalization and filtering methods are used.
The main objectives of this research work are as follows.

- Information on the Web is spread out across various mediums. The general framework presented in this thesis hides the complexities from users and gets data under a single platform using intelligent multi-agents.
- Along with making the process of finding information on the Web, the framework should utilize a hybrid approach of combing remote knowledge bases on the internet and local knowledge bases to enhance the quality and retrieval rates of data.
- To study the theory of different types of agents in order to short list essential key features of agent technology that can be used to tackle the research problem.
- Integration of third party tools using agent technology and then making them accessible in the system seamlessly.
- To study and document agriculture knowledge in the form of knowledge bases and to test the designed framework in the selected domain.
- To use agents to increase partnerships with universities and research institutes.
- To provide medium for information sharing and promotion of new technologies.
- To provide real time information of dynamically changing entities like (Weather, Market Prices, Market Trends, Soil testing locations ... etc).
- User friendly front end development is another objective to help gather personal information about users so their preferences can be stored in databases. This will help customize the pages according to their needs and wants.
- Fuzzy logic is used to handle imprecision and uncertainty found in Information captured in user profiles.
- Lot of valuable information is in research institute web pages. Often this information is not passed on to users in a timely manner. By getting all this information at one central portal it is easier for user to access it than to access each individual page. Not all users are looking for the same information. User’s role and their interests are pulled from databases by agents and these are used while presenting information to users.
- To facilitate communication with experts and other users in the same industry with the use of a centralized portal.
As stated earlier in the chapter, the overall aim of the thesis was to investigate how multi-agent system can be used to better handle information overload problem. The problem of information overload exists in many domains like tourism industry, health industry, education portals etc. Since the base framework is generic enough it can be applied to any domain to better handle the issue. The framework that is developed as part of this research work is used to tackle real world problems faced by farmers in India.

The design of multi-agent system was developed by taking cues from cognitive science. The methodology of experience, example and strategy fit the bill perfectly. Examples are a type of learning; they provide solutions when needed without much effort. Farmers can benefit enormously when these examples are bought to them in a timely manner. By knowing about these examples farmers can make more informed decisions. Just like examples experiences of other farmers provide valuable knowledge. Through this portal we provide a platform for farmers to share it with others. The final piece of the puzzle is strategically using intelligence to retrieve the right kind of content to present to users. All these elements are put together in the multi-agent system to provide a wonderful tool for farmers.

Along with using the principles of cognitive science a hybrid approach for data storage and retrieval is also devised. There are advantages to storing data in local knowledge bases but then maintenance will be a big issue. Similarly the Internet provides more up to date and realistic information but when resources are a constraint then it becomes a big hindrance to access that knowledge. This research proposes a way to combining the best of both. When a search is made data is first looked for in local database so retrieval time will be faster and data is more reliable and genuine. If data is not found in local database then search is extended to internet. If users find the data useful then by simply clicking a link it can be saved in local database for future searches. Along with the hybrid approach for data retrieval and storage appropriate use of customization techniques further enhance quality of search results. This research shows how all these methods can be implemented using multi-agent systems. Multi-agent systems provide a robust and flexible environment to easily integrate third party tools which will aid in getting timely and right content to users.
1.3 Technologies for the Research Work

At present, knowledge retrieval, especially knowledge retrieval based on Web information, is an important research subject [116]. A promising approach to information overload problem is to access large number of farming data sources through an intelligent multi-agent based framework. In this framework a set of agents cooperate with each other to retrieve relevant information from different sources. Multi-agent systems are best suited for solving complex real world problems. They have been successfully implemented to solve many practical problems in various domains like air traffic control, manufacturing, telecommunication systems and medical care.

The proposed solution is the development of effective and efficient autonomous software entities called intelligent information agents. These agents have the capability to access multiple, heterogeneous and geographically distributed information sources. A unique hybrid approach of getting content from local knowledge bases and web is proposed for taking advantages in both kinds of resources and handling their shortcomings effectively. Figure 1.3 presents a high level view of such a system. Users can use the system built on multi-agent system to view data that is customized, filtered and integrated in a single place. Integrating content from multiple sources saves time for users and further application of customization and filtering methods effectively handles information overload problem.

Figure 1.3: Proposed Architecture for Information Retrieval
Some of the key features of the system include personalized information retrieval, pro-activeness, ontology enhanced searches and fuzzy based rating system. The goal is realizing the power of data while minimizing the risk.

Modular design as depicted in Figure 1.4 works very well to meet the objectives presented in previous section. After a careful review of the objectives and the technologies selected for this research work this design is articulated. The aforementioned multi-agent system is enriched with the web based interface using agent technology.

![Figure 1.4: A Multi-agent Framework for Agricultural Activities](image)

The framework can be broadly classified into four main layers, front end layer, repository layer, agent layer and back end layer. Semantic web technologies are used in the front end layer to make the content more users friendly. MySQL is an open source free database that is used to hold local knowledge base and mapping information. Different kinds of agents are used in the system like profile agents, data retrieval agents, update agents, utility agents and monitoring agents. There are a number of third party tools which aid in information gathering and dissemination and agent technology provides the flexibility for easy integration into them. Back end layer looks into different tools that could be used. Each layer has specific functions they carry out and the detailed workings of them are given in later chapters.
1.4 Thesis Organization

The remainder of the thesis is organized as follows:

Chapter 2 goes over the concept of agent based systems. Detailed explanation of what an agent is and capabilities of multi-agent systems are covered. An overview of different agents and their roles is also presented. JADE is the middleware tool that is used to develop the proposed prototype. Justification for use of ontologies and fuzzy linguistics and how they can easily be used in agents are given in this chapter. Literature survey gauges the existing solutions for the research problem presented and their solutions.

Chapter 3 focuses on the general architecture of multi-agent system that was developed with agricultural activities in mind. The generic architecture consists of experience (implicit), strategy (explicit) and example (reasoning) and how these cognitive techniques are applied to build the prototype are explained in this chapter. The architecture consists of four main layers back end layer, agent layer, repository layer and front end layer. Along with coverage of the four areas the components that make up the areas are also explained. An in depth discussion of key technologies used and how the different agents collaborate and cooperate to optimize data storage and retrieval process is given.

Chapter 4 explains the detailed methodology of the generic architecture proposed in the previous chapter. The workings of multiple components and detailed explanation of four layers that handle key functionalities are covered. The way Ontologies and fuzzy linguistics are incorporated into the system is elucidated. Jena API is used to handle logistics of getting data from ontology models. SPARQL is the query language that is used to search Ontologies. Usage of JADE and Tomcat and how they are run from inside eclipse development platform is explained. Database design and creation of tables in MySQL database is also talked. Tools like Yahoo! Pipes, Google API and YouTube API are used in retrieving and handling information and integration details of them are also given in this chapter along with proper examples. To keep the system running smoothly a few utility and monitoring agents are used.

Chapter 5 discusses results of the framework and explains application features and their benefits. Real and valid data that is collected from multiple sources is loaded into local database tables. This helps to test all the functions and features in the system completely.
Chapter 6 looks into the research contribution made by this thesis in the field of computer science. This framework shows how intelligent agents can be used to pull data from local knowledge bases and also from Web. This helps improve response times when required data is available in local database and at the same time the application can surf Web if more data is required. Along with intelligence in content retrieval usage of customization methods and integration of third party tools help effectively tackle information overload problem.

1.5 Conclusion

This chapter gives an overview of the research topic considered. The background of existing scenario is discussed. Objectives that are going to be met as part of this thesis work and material that is going to be covered in the next chapters is also presented here. Based on the objective set, a broad framework to achieve the intended goals is also suggested.

The main goal of this research is to perspicaciously tackle information overload problem. Taking cues from cognitive science and using the robustness and flexibility of multi-agent systems a novel approach was designed.