THE FLORIDA AGRICULTURAL AND MECHANICAL UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

A FRAMEWORK FOR DESIGNING SMART ALUMNI SYSTEMS

by

LAKSHMI PRASANNA GANDHAM

A MASTER'S THESIS

submitted to the Faculty of the Computer and Information Sciences Department,
The College of Arts and Sciences and
The School of Graduate Studies and Research
In Partial Fulfillment of the Requirements
For the Degree of Master of Science in Computer Science

Fall 2011

The members of the Committee approve the thesis, entitled, "A Framework for Designing Smart Alumni Systems," by Lakshmi Prasanna Gandham, defended on December 1, 2011.

_____
Hongmei Chi, Ph.D.
Thesis Director

_____
Bobby Granville, Ph.D.
Committee Member

_____
Bhanu Prasad, Ph.D.
Committee Member

_____
Bruno Guerrieri, Ph.D.
Outside Committee Member

Approved:

_____
Edward L. Jones, Ph.D., Chair, Department of Computer and Information Sciences

_____
Ralph W. Turner, Ph.D., Dean, College of Arts and Sciences

_____
Verian Thomas, Ph.D., Interim Assistant Vice President for Research and Dean, Graduate Studies

ACKNOWLEDGMENTS

I would like to thank Dr. Hongmei Chi, my thesis advisor for guiding me throughout the master's program. Thank you for the many challenges and opportunities that you provided for me. Thanks are also due to my committee members: Dr. Bobby Granville, Dr. Bhanu Prasad and Dr. Bruno Guerrieri for their time and valuable feedback.

A special thanks is due to Dr. Edward Jones for all his support and mentoring during the course of my graduate study. Thanks to the Department of Computer and Information Sciences and Graduate School for financial assistance. Finally, I love and thank God, with whom all things are possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

A recent trend in society is social networking, a powerful tool for people to meet and interact based on common interests. Data mining is another powerful tool used to understand the vast amounts of data that are produced by social interactions, in order to enhance the services being provided, and for marketing. University alumni systems exist to promote active and ongoing relationships between graduates and their alma mater. This research proposes to incorporate selected features of social networking and data mining into alumni systems. Such an alumni system is termed a smart alumni system (SAS).

There are two major contributions of this research, a framework for smart alumni systems, and a proof-of-concept prototype implementation of an SAS subset. The SAS framework expands stakeholder roles beyond alumni to include current students, faculty, staff, and guests. The framework supports social networking style interactions within and across stakeholder types, for activities such as mentoring, fund-raising, curriculum development, etc. In the SAS framework, the primary purpose of data mining is to provide recommendations for establishing associations between stakeholders; a secondary purpose is analyzing results from university and departmental surveys.

The proof-of-concept smart alumni system prototype has been implemented as a web-based web system. The prototype implements stakeholder roles for students, faculty and alumni, and supports social networking features of friends, groups and messaging. Basic data mining algorithms are used to provide a ranked list of recommendations for stakeholder relationships for friends or groups.

1. INTRODUCTION

Alumni are one of the most important assets to any university. They are the people who represent the university in the real world. Many alumni networks were initially started from regional groups of alumni brought together for university fundraising activities. Later, these networks slowly gained added importance in the development of the universities because of their enormous outreach potential that benefits the university and helps current students in their career paths. The alumni groups have been in existence for decades and they are constantly changing with time. There have been very big changes in the recent years with the development of the internet and social networking that forces the alumni system to undergo huge changes. Therefore, it is really important for universities to focus on the alumni networks and find ways to enhance their growth and development.

A new intelligent system called ‘Smart Alumni System’ (SAS) proposed in this thesis offers an inter-mix of traditional alumni systems and social networking sites. A brief study on both systems is undertaken and the important features of these systems are incorporated into the new proposed system. Then the concept of data mining is used to pre-process large and complex datasets in extracting non-obvious patterns of correlations by removing unrelated data (noise) to discover hidden non-obvious patterns that represent valuable knowledge discoveries. It is a process of observing patterns in the data and summarizing the findings in terms of usable information. When carrying out the data mining operations on the system, some essential associations are revealed which should be studied in the development of more useful university alumni systems. This system

informs the university of the main areas to seek donations, improve interactions and determine from which parts of the world people are more interested to join the university. This thesis describes SAS and explains the important features and the results of data mining in detail.

1.1 Need for better alumni systems

What is social networking? This activity can be described as the act of sharing the associations of business or social relationships for the purpose of exploring common needs, interests or goals. Networking is and has been taking place all over the world since the existence of humans. Every person has a network; some recognize it and some do not. Nowadays networking has become easy because of the development of technology. People can interact with one another, share the ideas, and discuss their personal and professional lives without even moving from their desks. Existing alumni systems are usually developed to facilitate networking between the alumni and their respective universities, but most of these current systems are not being used by the majority of the alumni for many reasons.

Figure 1: Functions of the current alumni systems

A simple glance at most of the current alumni systems can give us a few answers. Typical alumni systems are static, not engaging and are usually confined to just having information of the universities, upcoming activities and they have very little to no space for interaction among its members. Traditionally such systems are accessible only to the alumni, and do not involve current students and faculty, thereby hampering the realization of numerous potential benefits. As such, there is a need for systems, which bring on board alumni, current students, staff and faculty, to facilitate constructive interactions which could contribute to building a stronger university legacy.

As can be seen in Figure 1, the current alumni systems do not provide any channels for interaction, especially with the current student community. A few preset channels for interaction are available and they are unidirectional from alumni to university or vice versa. The alumni systems should be developed in ways they increase the interactions

among various categories of people associated with the university, like between current students and alumni, between alumni and faculty, between current students and prospective students or their parents. Such a framework would place current students within reach of the expert advice of alumni and of their stories on how to overcome real challenges in the real world. Prospective students, on the other hand, could use it to learn more about courses, college routine from current students. Such systems could also cater to the needs of the parents, who wanted to know the about the universities from the current students, alumni and also the faculty. As the current alumni systems do not have any of these features, there is a serious need for re-engineering the alumni systems.

1.2 The Concept of Smart Alumni System

The Smart Alumni System is a combination of the traditional alumni system and social networking sites. Unlike traditional systems, SAS will be interactive in nature and have room for multiple levels of interaction among different categories of university related personnel, and will focus more on the common goals of the involved personnel or the university itself.

The challenges faced by current alumni systems are:

- to attract alumni to access system like other social networks;
- to collect useful data related to alumni from social networks and other open resources;
- to keep privacy and security a top priority for this system; and
- to get current students involved.

All of the above challenges can be addressed through a Smart Alumni System, which is not just confined to the alumni of the universities. There are different stake-holders in the system, ranging from the alumni who have studied in the university to visitors or prospective students who want to know more about the university. All members in the system will be divided into four main categories: university, alumni, current students, faculty and visitors along with the administrators. Each of the members will have some features, which are specific to their individual group (of users) and some general features which tend to be common to all users.

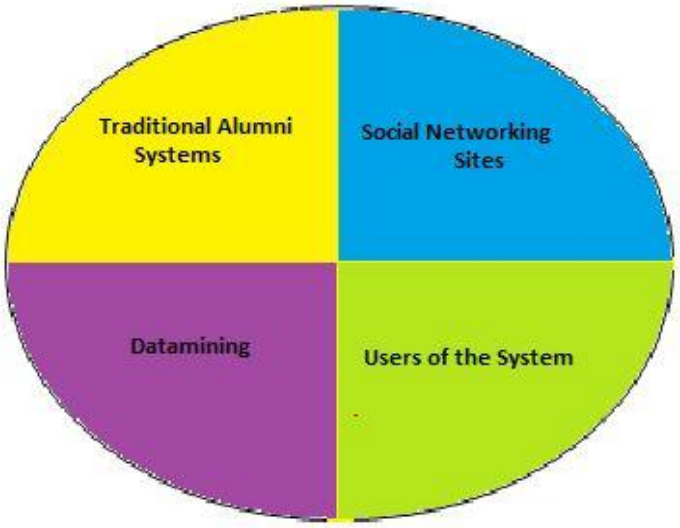

Figure 2: Components of the Smart Alumni System

Typical examples of common features required for all of the members in the groups include maintenance of individual information profiles and privacy aspects, which allow selective sharing of information intended to be shared. Each member can share his or her thoughts and other information using SAS as the medium. Also, each member can

communicate with other members in the system, subject to some restrictions inherent to the type of the user. The other important features, operation of SAS and the advantages of having this system will be discussed in more detail in the upcoming chapters.

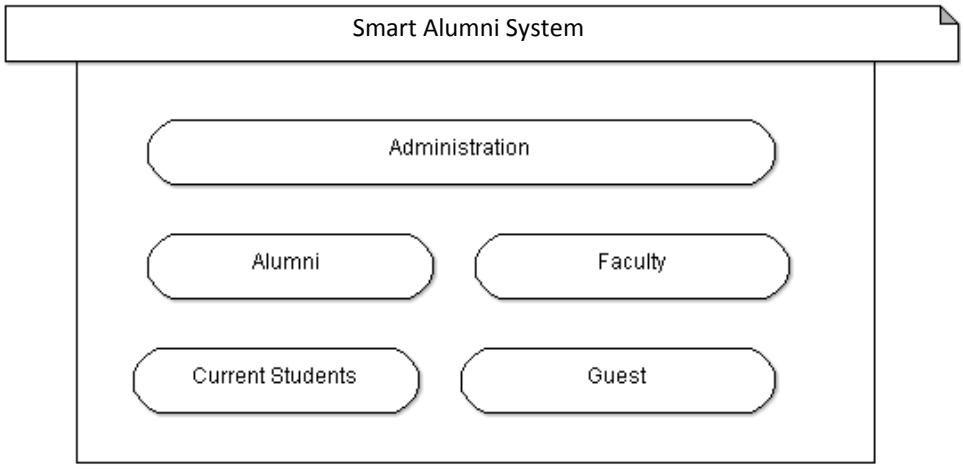

Figure 3: Stakeholders of Smart Alumni System

One of the most distinguishing aspects of the SAS is the usage of data mining principles to discover additional relationships among the users. Data collected from alumni systems is analyzed for any possible trends (or patterns) characteristic of the observable, and harnessing the information to help tune the development of the system. The trends can be of various kinds like assessing student interest in joining the university based on their geographic location of origin, or assessing how well a mathematics major fares with computers compared to students in science majors, or assessing which group of alumni tends to pledge money for the university. All such assessments can be performed using the data collected on the SAS site; they in turn would help the university to determine the areas of focus relevant for different aspects for future development.

For example, suppose the university wants to raise funds and is not sure which group of people would respond when contacted, data mining techniques can be used on available data to answer this question. As shown in Table 1, there is a dataset of alumni, indicating their willingness to give donations to the university. The dataset has four attributes and one outcome.

Table 1: Example of data collected from alumni system: Donor Attributes

Alumni Year	**Current City**	**Highest Degree**	**Occupation**	**Pledge money**
1986	Tallahassee	Masters	Software	Yes
1987	Orlando	Masters	Entrepreneur	Yes
1986	Tampa	Undergrad	Software	No
1990	Tallahassee	Masters	Govt. Ser.	Yes
1985	Miami	Undergrad	Software	Yes
1986	Tampa	Undergrad	Govt. Ser.	Yes
1985	Orlando	Undergrad	Software	No
1984	Orlando	Undergrad	Software	No

This data creates $5 \times 4 \times 2 \times 3 = 120$ combinations out of which 8 are shown here. Now using various data mining algorithms one can extract some rules like below:

- If "Alumni Year" is 1986 and the highest degree is "Masters", then – "Pledging money" is a "yes";
- If "Alumni Year" is 1990 and the "Occupation" is "Govt. Ser", then – "Pledging money" is a "yes"; and
- If "Alumni Year" is 1984, then – "Pledging money" is a "no".

This illustrates a very simple process of extracting human-understandable structure or patterns to determine which alumni group of students are willing to donate money to the university. More detailed use of data mining and it advantages are detailed in later chapters of this thesis.

1.3 Motivation and Challenges

Many challenges exist for building the desired alumni systems. Most people are habituated to social networking sites and spending time on these sites has become an indispensible activity in the daily routine. As such, it may be quite a difficult task to make alumni and current students to get habituated to SAS. A factor which could attract more members is the current member database. If more users are currently using the system, it is more likely that it would attract new users. Further, if the current students are allowed to be a part of the SAS while enrolled at the university, the university need not put any additional effort in attracting them or getting them involved with alumni activities once they graduate. A simple change of status from current students to alumni can be performed which correspondingly changes the privileges of the particular member.

Several issues need to be considered while expanding the alumni system to include current students. The principal logistics to be considered are: the hardware to support larger websites; simultaneous handling of privacy and alumni concerns. It should be kept in mind that not all alumni might be interested in sharing their details with people whom they don't know. The increase in data volumes resulting from larger website system also adds to questions about the privacy of the data. This can be taken care of by specifying selective permissions to members in the system. Based on the type of user the access can

be defined. One key aspect in protecting privacy can be accomplished by requiring all the members in the system to have a valid email id, provided by the university. The guest users would have an exception to this rule, and will be allotted restricted access. Another rule which can be incorporated is the validation of information provided by a user during registration to check for consistency. This information typically involves details such as student number, department, major, year of graduation etc and a combination of these values are unique. Hence, only authenticated members can get access to the system.

1.4 Related Work

The need for maintaining alumni systems has been highlighted by Arceo [25] in his study on the role of student and alumni associations in democratization of Spain during the post-Franco political transition. Arceo explained how student and alumni associations of Spain served as a bridge or boundary spanners between the educational institutions, their people (students, alumni, professors and political sector) and the job market. Eventually, these associations grew in their presence to decision-making bodies which helped retraining of the unemployed youth and thereby contributed towards development of people's lives through decreased unemployment.

One of the early efforts to establish an interactive alumni association is due to Spasić and Pejak [26] at Faculty of Mechanical Engineering, University of Belgrade. Based on experience with previous alumni association Spasić and Pejak established an interactive alumni system called ALFaMEB or αMEβ, which would cater to quality assurance in higher education. Feedback information was collected through various activities as well as questionnaires about industrial expectations; knowledge obtained at university, and

need for improvement in teaching and research activities. The alumni ALFaMEB activities helped continuously improve the teaching and research undertaken at the university and thereby broaden the limits of university education.

A noteworthy work in the present context is by Barnard [27] who proposed an online community portal which is beneficial to both alumni and higher education institutions in South Africa. The portal was a web-based secured database which allows free sharing of information between alumni management teams and its members. Using empirical survey methods, the amount of information disseminated to target alumni audiences at the University of Johannesburg and their opinions on community portals is collected. Based on the collected information the needs for building a network which is beneficial to both the alumni and the university are identified and a prototype portal is built which discards the 'on-size-fits-all' notion of information sharing among alumni. However, this study was limited to alumni members and university as its stakeholders and had provision for limited interactions between the alumni.

Delavari et al. [28] investigated the capabilities of data mining when applied in the context of higher education system. The authors propose a roadmap analysis model DM_EDU which is aimed at improving the decision making procedures related to various processes which are essential to maintain the quality of educational system. Knowledge discovered by application of data mining techniques to raw educational data assists decision makers to improve the decision-making procedures and to set more enhanced policies for educational processes. Also, by understanding of student enrollment tendencies and patterns in a course, they were able to predict if a student is likely to

perform well in a given course or otherwise. Such predictions also help in identifying necessary actions required to provide additional skill classes and attention to the students who are likely to fare below par. As such, it was shown that decision making processes such as planning, counseling and evaluation can benefit from application of data mining techniques to educational data.

Although the work discussed above makes novel attempts to improve student or alumni based networks systems, they are severely limited in their communication capabilities and fall short of harnessing the full potential of extensive interactions between student, alumni and university. This thesis aims to address all those shortcomings by providing a prototype 'Smart Alumni System' (SAS) which has all the desired features of interactive alumni networks and is supported by decision making processes based on data mining principles.

1.5 Thesis Organization

The organization of this thesis is as follows. Following this introduction, Chapter 2 discusses about alumni systems and how they have transformed over the years. Chapter 3 presents a topical introduction to data mining and explains its basic use. Chapter 4 focuses more on fusion of data mining concepts with alumni systems. Chapter 5 introduces the design of SAS and chapter 6 discuses the implementation of the SAS. Finally key conclusions and future work are summarized in Chapter 7.

1.6 Contribution to Research

The main objective of this thesis is to develop a framework for an interactive-social alumni system. The following are the main aspects of this thesis in that regard.

- Development of a framework for Smart Alumni System (SAS), and its implementation. While studying feasibility of such systems focus is given to rules on stakeholders of the system, their attributes, type of interactions, functions, accessibility restrictions and benefits due to SAS.
- Incorporation of social networking features in the Smart Alumni System. Focus is given to identification of key social networking features, their using as a part of the proposed SAS.
- Application of data mining principles to extract useful information from the data collected as a part of SAS. Based on the type of user and his desired interaction with other SAS members, the user is recommended of appropriate suggestions obtained through data mining.

2. ALUMNI SYSTEMS

Alumni systems are web-based information systems which are created by the universities to maintain relations with its students after they graduate. Universities use the systems to collect alumni information and organize outreach activities. Typical systems collect information about the current location, workplace, job information and even hobbies of the alumni.

2.1 Necessity for alumni systems

Many factors necessitate the alumni systems. Several studies show that having good relations with the alumni helps in the development of the universities [1, 2]. In view of many potential benefits, some universities initially set up alumni systems as a means to collect funds from the alumni. The recent recession has led to severe cuts in government support for development and educational activities at most universities. As a result, universities and educational institutions need to secure alternative sources of cash inflow. It is beneficial to the university to maintain healthy relationships with the alumni who have the financial solvency to donate to their home university. Hence, universities all over the United States and the world started exploring alumni databases and relations with the intention to obtain monetary gifts.

Another important factor which necessitates alumni systems is the university's need to market itself as a leading place for education and to attract new students. Given the tremendous amount of competition in attracting good caliber students from around the world, universities are competing to stand out as leading educational institutions that produce quality professionals and successful pioneers. Most universities tend to boast

about their alumni's research works or achievements as one of their own. In some cases, universities also tend to invite famous and successful alumni to given guest-lectures or talks to motivate current students. This also helps the universities to advertise their affiliation with successful alumni. Another popular way to market the university as an excellent educational institution is to show positive employment statistics of its graduates. It is not surprising that some of the high profiled universities have most structured graduate follow-up programs, where employment of their alumni and their achievements are tracked continuously.

The two factors — relationships with alumni and marketing — can be seen as some of the many reasons to initiate alumni systems. With changing times, the perception of universities with regards to the use and scope of alumni systems has also evolved. It is envisioned that such alumni systems are beneficial not only to the universities but also to alumni and current students. It could serve as a platform for current students to interact with their senior alumni and receive mentoring from them with regards to career paths, real world expectations and so on. From the alumni point of view, benefits of participating in alumni groups allow them to stay in contact and meet their fellow classmates, professors and university personnel, even after several years beyond graduation. It also allows alumni to share professional or personal advice, explore mutual interests and exciting new opportunities in a collaborative way.

2.2 Traditional Alumni Systems

Prior to the internet revolution, universities had limited ways to engage their former students. Communications were through postal mail, letters or magazines and since the

alumni are likely to be geographically scattered in search of their own careers, the universities' efforts to reach alumni was affected by publication and mailing costs. Most communications were informative in nature and not interactive. Alumni interest and participation in university news or its affairs was usually higher in individuals who were stationed within a geographic proximity to the university, or in people whose family or friends still study at the university.

Following the internet revolution, though official alumni systems were put in place, most of them were an add-on to university websites and not a lot of alumni would get involved with it. These early alumni systems suffered from several drawbacks. The focus of early alumni systems was mostly for collecting university funds and due it its rudimentary structure offered little to no channels for interactive communication. In addition, the number of alumni actually participating or retaining university relations was only a fraction of the total alumni population due to limited access or technological means to use alumni systems.

2.3 Present Trend in Alumni Systems

Recent developments in information technology and internet have resulted in faster, easier and less expensive means of communication. Universities have well maintained websites dedicated to university news, departmental, course and faculty information. Using the same means, most universities developed alumni information systems which typically list the information of student alumni like their year of graduation, current occupation and other details. Such systems, while providing a good listing of alumni details through dedicated pages on university website do not have any medium to

facilitate interactions between the alumni. The main theme of such systems is to have a listing of alumni contact information which is used for the purpose of actively pursuing donations.

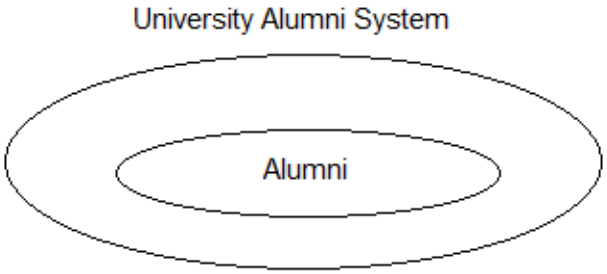

Figure 4: Traditional alumni system structure

Some universities have developed attractive alumni websites, but the major drawback of these systems is that they are limited only to alumni.

2.3.1 Alumni Systems in Social Networking sites

The most recent development in alumni systems is the use of the social networking sites to connect with alumni [3]. Nowadays, there are many third party online systems that host alumni systems. Almost all students have an account in a social networking site such as Facebook, Twitter, LinkedIn, MySpace; these sites have an open user approach where anyone can join and co-opt the online activities of the alumni. Third party sites are playing an important role in alumni organizations by providing a convenient platform for students to communicate. Some universities are exploring this advantage and trying to make use of social networking sites to maintain its alumni systems instead of creating one

of their own. Universities with existing systems are trying to link their alumni systems to existing professional and personal networking websites.

The social network sites have also gone a bit further in catering to alumni systems. For example, the Caltech Alumni on LinkedIn have a Caltech Alumni Logo which appears on their profile next to their name. Such an identification tag allows other Caltech alumni to advertize their own connection to Caltech and aids in having contact settings which allow easy communication with other alumni. The link between the alumni and non-alumni network acts as a multiplier that magnifies the value of each network [4].

2.3.2 Role of Twitter/Face book in Alumni Networks.

2.3.2.1 Facebook

Facebook has become a very powerful social networking site in the recent years. Alumni networks using facebook can be developed in two ways: (a) by creating a fan page for the university; (b) by creating an exclusive group within the facebook system. Both of these options allow their members to connect through the group or fan page and provide a platform for discussion. Once a Facebook user finds the fan page a simple click on the “like” tab, establishes a link between the user and the fan page of the university. The user need not explicitly check for university news or alumni related updates everyday as the updates if any would be automatically posted onto the wall of each member. In the case of group pages, the user can join them directly if it is open for public, or may need to get approval from an administrator, in case the group has a closed or moderated setting. Universities and even departments are creating these pages and

groups in facebook to advertise about themselves, the courses offered, sports related activities and various other activities where the alumni along with other interested people tend to join.

2.3.2.2 Twitter

Twitter is a networking tool based on micro blogging and can offer a unique option to educational institutions who want to establish connection between the institution, its alumni and friends. Twitter may help fulfill two primary tasks: (a) delivery of information and (b) increase in active alumni user population. Universities or its dedicated alumni organizations can use twitter to deliver up to date news and information about their institution, current students, faculty and alumni. Since twitter is a free service, any alumnus can join with ease and alumni associations therein can inform its current users of the new members and encourage them to follow the new users.

The use of social networking sites as medium for an alumni system might not be free of some disadvantages. The design of social networking sites might raise severe concerns on the privacy of individuals and possible misuse of information provided on the social networking sites. In addition, these sites might offer a few less interactive channels of communication might achieve only selected focus of university with regards to fund raising or advertising.

2.4 Need for Smart Alumni System

Most of the alumni systems to date are focused just on university alumni. However, having just the alumni in the system will not be of much use either to the alumni or the university. One could expand the member base to include current students, professors,

other affiliated staff members and guests or even people who are curious to know about the university. Such an expanded heterogeneous system tends to facilitate multiple kinds of interaction between its constituent members and can help to bring aboard all the people associated with the university on a single platform to discuss everything from upcoming activities, news and research to career programs. This is precisely the motto behind the proposed smart alumni network. The main stakeholders in the system are alumni, current students, faculty, administration, staff and guests. Given the different types of users, the Smart Alumni System must provide efficient communication channels. This can be accomplished by offering social networking feature as part of the alumni system. Members in the system can virtually interact one-on-one or in a group, and the user data or data resulting from their interaction can be used to extract useful information for further studies or use, through the application of data mining technologies.

2.5 Advantages of Smart Alumni System

Smart Alumni System works better than the current alumni systems in various ways. The most outstanding feature of SAS is it permits students to join before they graduate from the universities in this way it will not be difficult to encourage them use the alumni system after they graduate. SAS provides a ‘single point of network’ as they connect all the people connected to the university at a single place allowing interactions, exchange of ideas and other information.

2.5.1 Social Networking Feature

The main advantage of the SAS over traditional systems is the empowerment of its users to accomplish 'networking tasks'. Each full member can enter into the system through an exclusive log in and virtually meet other members in the system. The users will be able to have profiles setup in the system and can share information about themselves, their education, career, interests and everything they want to share in with the other users in the system in an exclusive network.

2.5.2 Benefits to the University

Universities are one of the main stakeholders who get benefits from the alumni system. Some of the main benefits are:

- *Fund raising*. Universities can organize events for the alumni, which in turn act as fund raising activities for the universities;

- *Marketing*. Universities whose alumni are reputed and notable can promote the brand of the university through highlighting their affiliation with the noted individual;

- *Improving standards*. Departments within a university can come along with the alumni and discuss on the new emerging technologies in the real world and make necessary changes to curriculum. Consequently increasing the quality of education obtained by the students.

2.5.3 Benefits to the Alumni

In addition to the university, the alumni systems also prove to be advantageous for the alumni. Each alumnus is likely to have a unique life experience and alumni systems provide an ideal place for them to share their stories and life experiences. It can also facilitate re-unions between old friends and contribute to their relaxation and health. Alumni who turned into entrepreneurs may benefit by recruiting fresh graduates from their home university, a place whose educational standards they trust. Some of the advantages include:

- *Exclusive networking*. Online alumni networks are very exclusive social networking environments. These are very good platforms for alumni to meet their class mates and fellow students and can maintain continual contact with them in future.
- *Career path assistance*. Alumni get together through socials or by online means. Interactions between alumni may result in valuable career path change if necessary.
- *Special benefits*. University provides special benefits for their alumni, like providing game tickets, university email, special parking permissions etc.

2.5.4 Benefits to Current Students

The traditional alumni networks provide most of the benefits to university. SAS would bring together currently enrolled students along with the alumni thereby leading to many potential benefits.

- *Mentoring*. Current students can choose an alumni mentor with whom they can interact and take advice from.
- *Recruitment*. Current students get to know about the job openings in the workplaces where the alumni are working and in some cases the alumni can even act as a reference provider.
- *Beneficial Interactions*. Current students can interact with the alumni who are experienced in different trades to get updated about the trending technologies and opportunities in the technical world. Students who relocate to a new place can use alumni systems to make connections with the alumni living in the same area.

2.5.5 Benefits to the Faculty

SAS offers numerous benefits to faculty members in the system:

- *Good means to interact with students*. SAS provides lot of channels for the members to interact, the faculty can use it to express their views and share interests.

- *Obtaining feedback*. SAS is not like traditional alumni systems, and provides a good opportunity for the faculty to obtain feedback about their teaching or the courses content, and if required they can try to improve it from a student perspective.

2.5.6 Benefits to Guest User

The guest users in the SAS include staff, prospective students, parents and potential recruiters. Each of these users uniquely contributes to the overall user experience of SAS through improved information exchange. For example, the staff can assist with

department operational questions; or assist students seeking help with various jobs involving logistic management. The prospective students and parents can have direct access to university personnel, staff and faculty, to whom they can directly ask their questions or concerns and obtain additional information regarding courses, programs, and future prospects. Recruiters can also benefit from SAS as it provides them with a feel of the training underwent by the graduates of the university and helps in selecting individuals who match their company requirements the best.

Various levels of interaction between the different kinds of stakeholders in SAS provide them with unique benefits and helps serve the needs of other stakeholders as well. The interactions between different stakeholders in SAS are shown in Figure 5. In addition Table 2, summarizes the classification of stakeholders in SAS.

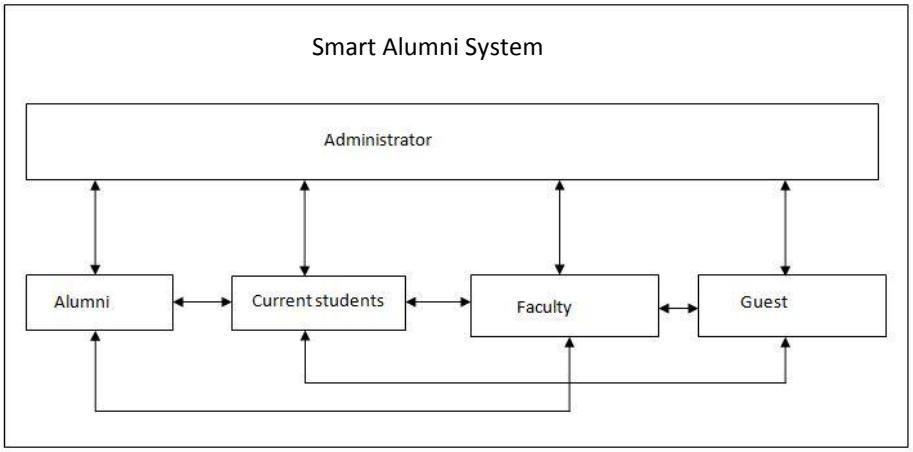

Figure 5: Interactions between stakeholders in SAS

All the stakeholders in the systems do not have identical access to the SAS. As shown in Figure 6, there are three types of access given to the users: full access, partial

access, restricted access. Full access users include university administrators or officials who have the privilege to setup the access to all other users of the system.

Table 2: Stakeholders of Smart Alumni System

Major Stakeholders	Subdivision	Description
SAS Administrators		People responsible for maintaining the system, giving access to users etc.
University		High level users
Faculty	Department faculty	Faculty belonging to the department.
	Visiting Faculty	Professors from the other departments or companies
Alumni Students	Undergrad	Students from the department
	Masters	
	Doctorate	
Current Students	From the department	Students from the department
	Out of the department	Students who take courses in the particular department
Staff	Technical	People working for the department in various sections
	Non-Technical	
Guest	Parents	People who want to know about the department
	Prospective Students	
	Recruiting Companies	

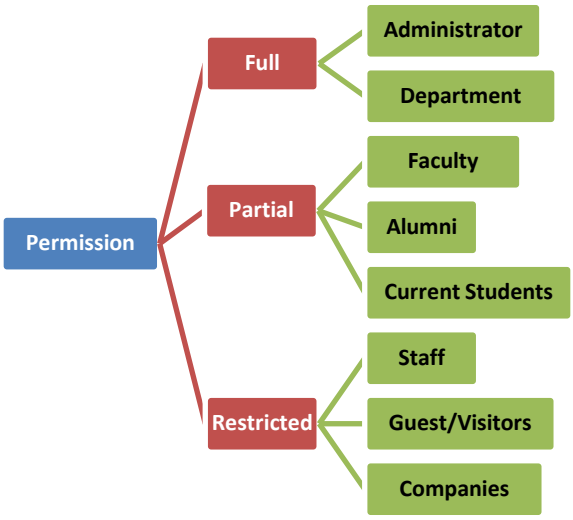

Figure 6: Type of access of different stakeholders

Partial access users include alumni, students and the faculty. They form the main clientele of SAS and have most privileges in the system sans any responsibilities involved in granting access to other people. Restricted access users include all the staff and guests whose features are very restricted in role. For example, restricted users may not be able to see the personal details in the profiles of member in the system. The three access levels are devised to ensure privacy of the user base.

3. DATA MINING

Data mining is the process of extracting patterns from data. It is becoming an increasingly important tool to extract meaningful information from data in a human-understandable structure [5]. Through appropriate use of data mining algorithms SAS can be used to harness volumes of data which can be transformed into information useful for the purpose of serving university or other students. This chapter provides a brief introduction to key data mining principles, its features, applications, advantages and privacy issues.

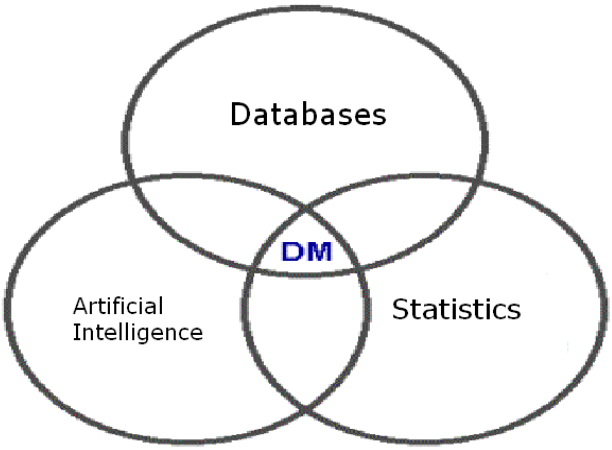

Figure 7: Intersection of technologies yielding to data mining

3.1 Introduction of Data mining

Explosive progress in networking, storage, and processor technologies has led to the creation of ultra large databases that record unprecedented amount of transitional information. All of this available information that can be gathered through the internet

has enabled many organizations to accumulate large amounts of data, most of which is probably not of much use. However, if ‘meaningful’ information can be extracted or filtered from such information heaps, it can help us learn about the subjects on whom the data are based and provide benefits to the collector of such information. Information is considered an asset that gives an organization a competitive advantage over its competitors and if used efficiently it can bring in high profits.

Traditionally statistical investigations based on datasets required analyzing heaps of data, a process which requires a lot of time. In recent years, data mining has emerged as a significant technology for gaining knowledge/meaningful information from the vast quantities of data without any prior hypothesis. One can view data mining as a technology which is basically at the intersection of three fields: database systems, artificial Intelligence and statistics. It has introduced new concepts and algorithms, applied machine-learning and is used in business and research as a popular alternative to make well informed decisions.

3.2 Working of Data mining

One could say data mining works the same way a human being does. It uses the historical information (experience) to learn. In order for the data mining technology to extract a useful pattern out of a database, one needs to identify how it would look like. The description of that information is used to look for similar examples in the database and develops a predictive model for future based on the available pieces of information [6]. The five major elements of data mining can be summarized as:

- Extract, transform, and load transaction data onto the data warehouse system.

- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in a useful format, such as a graph or table.

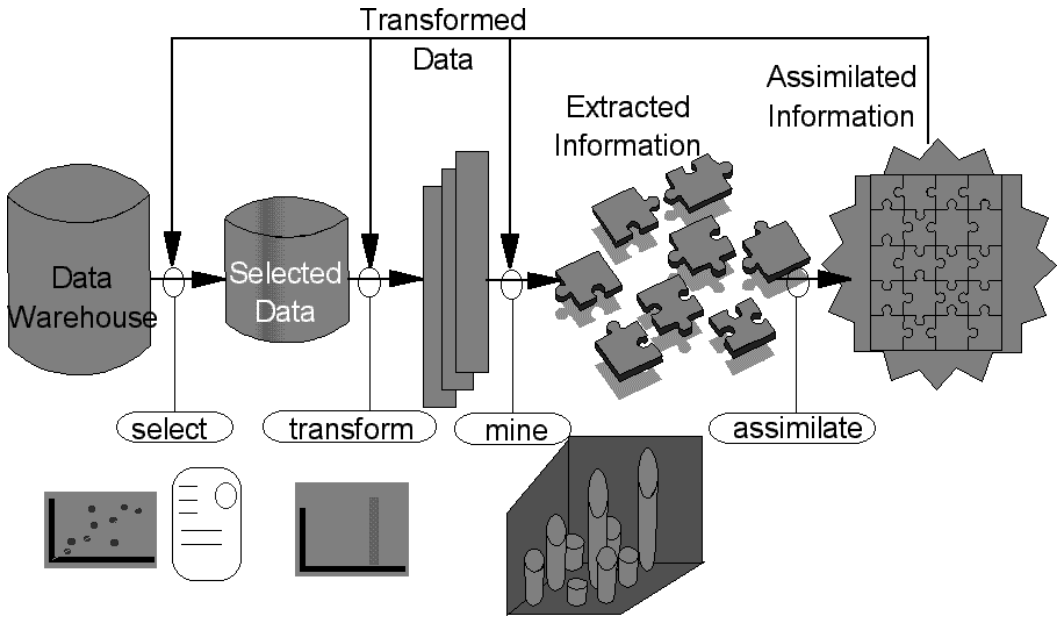

Figure 8: Schematic of the five major elements of data mining process [7]

3.3 Advantages of Data mining

Transforming raw data into information has many advantages in different kinds of fields. Data mining is about extracting essential information; it is the acquisition of relevant knowledge that can allow you to make strategic decisions which will allow the

business or organization to succeed. The following examples discuss the advantages of data mining in various fields.

Marketing and Retailing. Data mining can help marketers understand their customers purchasing behaviors. Marketers and retail outlets can design their advertizing strategies to attract and fit the needs of the customer. For example, a toy manufacturer can understand the purchasing trends of parents with children of different age group and customize their strategies to attract the consumer attention. Data mining can also help in making decisions related to development of new products, based on sales patterns of their previous products. Likewise, retail stores can also benefit from data mining[14]. For example, through the trends provided by data mining, store managers can arrange shelves, stock certain items, or provide discounts which are likely to attract consumer attraction.

Entertainment. Based on movies and TV shows watched by the viewers in the past, datamining algorithms can be used to predict the preferences and advice viewers of movies or programs they might like. For example, Netflix, an on-demand internet streaming media provider, tracks the movie choices made by users to provide them with suggestions among the available list of programs.

Banking. Data mining can help financial institutions to perform risk evaluation on aspects related to lending loans, credit cards, and investments. For example, by examining creditability of customers, a bank can estimate the level of risk associated in issuing a new loan. Though the assessments might not be totally accurate, it is found to help reduce the losses due to bad customers [8].

Medical Sciences. Data mining is being used in medical sciences helping doctors to understand the human genetics and bioinformatics. For example, mapping relationships between human DNA sequences and susceptibility to common diseases can be obtained. This would help doctors understand how variations in an individual's DNA sequence affect the risk of developing common diseases such as cancer.

3.3.1 Disadvantages of Data mining

Despite its potential benefits, the realization of datamining in practice has some potential drawbacks mostly due to concerns of security, privacy and information misuse.

Security issues: Most companies which collect personal information through their online portals might not always have sufficient security systems in place to protect it. As such, data collected might be prone to security breach. For example, Ford Motor Credit Company had to inform 13,000 of the consumers that their personal information including social security number, address, account number and payment history were accessed by hackers who broke into a database belonging to the Experian credit reporting agency [11]. Such incidents expose the potential vulnerability of sharing personal information, and possible identity theft problems.

Privacy Issues: Personal privacy has always been a major concern. With widespread use of internet in recent times, concerns about privacy of individual have grown. Some people tend to limit their online shopping to avoid any privacy breach. Though against the law to sell or trade personal information, such instances have occurred in the past. For example, according to Washington Post, in 1998, CVS had sold their patient’s prescription purchases to a different company [13]. The financial giant

American Express has also sold its customer's credit card purchase information. Such actions without the consent of the customers violates privacy laws and may also inflict harm as the seekers of such information can misuse it.

Misuse of information: Trends obtained through data mining intended for a specific marketing purpose may be unethically misused for other purposes. Practitioners of data mining principles, especially those handling people data need to act responsibly. There is a risk these practitioners might use the information obtained through data mining to discriminate against a certain group of people or even take advantage of vulnerable people. Since, the datamining techniques are not error free, any errors or misinterpretation can have serious consequence [12].

4. DATA MINING FOR SMART ALUMNI SYSTEM

This chapter introduces principles of data mining and demonstrates the use of data mining to improve the quality of Smart Alumni System.

4.1 Need for Data mining for SAS

Identification of information needed to improve a system is instrumental to the success and sustainability of the system itself. As in the case of any information system, the SAS is likely to accumulate volumes of data which need to be processed to extract strategically important information. Improvements in methods used for processing the data would translate to improvements in extracting strategic information and will lead to the significant improvement the system quality itself.

4.2 Identification and Design of Key Processes

The key to the success, quality and efficiency of SAS lies in the identification of major interactions/processes that happen among different stakeholders in the system. These processes play a key role and can also be further improved though data mining. Some of the key processes for different stakeholders are identified as below:

Alumni processes: The processes centric to the alumni users of SAS include identification of patterns among former graduates based on various factors such as their current occupation, job, pursuit of advanced studies; patterns in donations received from alumni; patterns in alumni willingness to mentor current students; and even patterns in community participation and activities of the alumni.

Student processes: The processes resulting for current student interaction with SAS may include identification of patterns classifying the behavior of male versus female students; indentifying amount activeness of students, any specific patterns in the nature of job or career path chosen by the students.

University processes: The processes which are required for university related interaction with SAS include patterns on identification of alumni who can contribute to university development, identification or donors, prospective recruiters; and identification of alumni affiliates which could promote the university itself.

Faculty processes: The faculty users in SAS can also yield in identifiable patterns which are of use. These include patterns among the faculty within a department, or inter-department; patterns among faculty with similar research interests; and patterns among faculty interaction with student and its role in the student's success.

Guest processes: The guest users in SAS may also yield patterns useful for strategic choices by the university. Some examples include patterns as a result of interaction between current and prospective students, pattern due to parent interaction with university personnel and patterns based on demographics of guest user and the chances of the user being a part of the system in future.

All these patterns can be recognized from the system by using appropriate algorithms. The following sub-sections describe the design of a few salient interactions in SAS.

4.2.1 Curriculum Development

Curriculum development is one of the important features of SAS. New technologies emerge continually and students need an enough exposure to emerging technologies. As such, there is a necessity for frequent changes and new additions in the coursework taught at the universities in order to prepare the students better for a competitive world market. Alumni working in industry can serve as best sources to obtain inputs for curriculum development. Typically, when the university intends to change or modify the existing curriculum, the professors who have instructed or researched in the pertinent topics are consulted for advice. In addition, if the alumni who are working in a select technology/sector are also asked to participate in the discussion activity along with the faculty, there are better chances for having a well-craft curriculum which is targeted to the currents needs of the industry.

The key, however, is setting up the mechanism to achieve this by identifying the alumni-faculty combination which can best provide with valuable suggestions and play a key role in the development. Such information, can be harnessed from SAS data through appropriate data mining tools. By design, SAS database comprises data relevant to both the alumni and the faculty. So if a simple clustering function is applied on the data, there will be people grouped under various clusters depending on the data attributes. In the specific case discussed above, the data attributes used to select the alumni may be their occupation, previous work experience, and current city; for the faculty data attributes like their department, teaching experience, and research interests may be used. When the clustering operation is performed on the data, one can obtain all the data partitioned into

various clusters and depending on the further necessity data specific to a cluster can be analyzed and used accordingly.

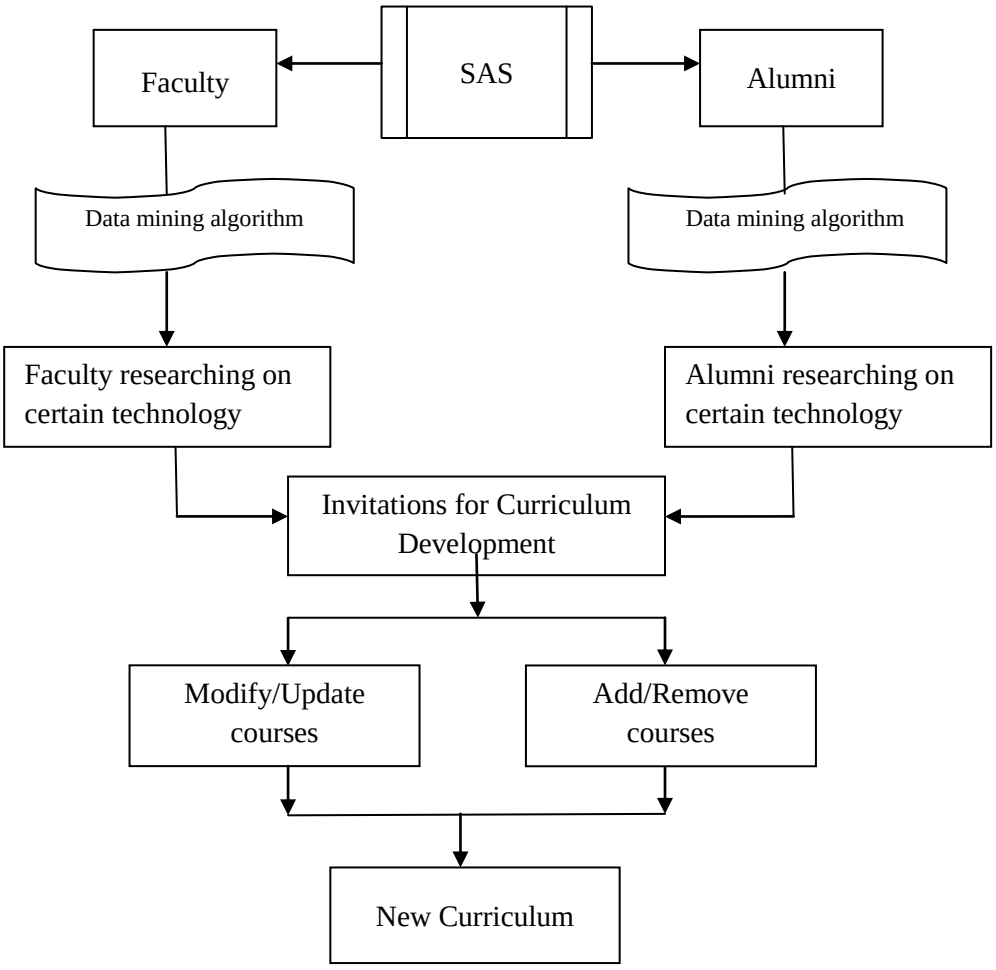

Figure 9: Schematic of interactions for curriculum development

Figure 9, shows the schematic of design for curriculum development feature in the SAS. The data mining algorithm (clustering) will be applied to each category of individuals and the data corresponding to faculty and alumni matching the requirement are identified and contacted with invitations to participate in the curriculum development.

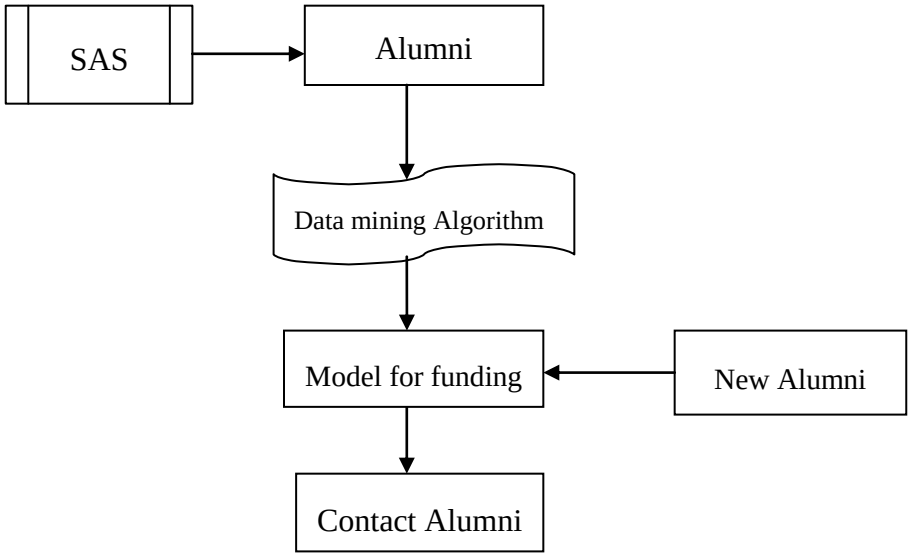

Figure 10: Schematic of interactions to secure funding process

4.2.2 Fund Raising

Owing to the recession in recent times, there have been cuts in funds for the universities and it has become very hard for some universities to survive. At this point, it is really important for the universities to search for other options for securing the required operating and developmental funds. The alumni can be perceived as one of the major contributors to the universities, over the long run. The SAS system has the information of all the alumni who are members of the system. All this data can be analyzed using the data mining functions to find out who among the alumni will be willing to contribute for the university as shown in Figure 10. Every detail of the alumni can be used from the place of residence, current position, and year of graduation etc. for finding out the likely patterns which are not known a priori. Once a model is developed it can be probed to further know the list of alumni who can be contacted.

4.2.3 Mentoring

Mentoring is a process where the alumni can guide the current students to make appropriate career choices. The alumnus who are typically working in a specific industry they have first hand information related to the requirements of securing a job in the selected field, technical and professional expectations at work, and likely career path options. The alumni can act as mentors for current students, helping them in many ways, mostly by explaining out career options and expectations. Alumni with interests and time availability to mentor a current student may express it in their SAS profile.

Alumni data can be searched for three types of patterns: alumni not interested in mentoring activities, alumni undecided about their mentoring interested and alumni interested in mentoring as shown in Figure 11. The first category of alumni need not be contacted. For the second category a communication can be sent to further encourage them for participating in the mentoring activities. For the alumni who have expressed interest and availability to mentor, their profiles are matched with the profiles of the students based on various attributes/fields like interests, occupation, current city etc. Data mining routines allows a better matching of the alumni with the current student profiles and tends to make the mentoring feature more fruitful.

4.2.4 Recruiting

Upon graduation students settle in various industries either related to their area of study of otherwise. The alumni can be perceived as a representation of the university at their work areas because the companies gauge the standard of education being taught in a particular university through their work performance. When fresh batch of graduating

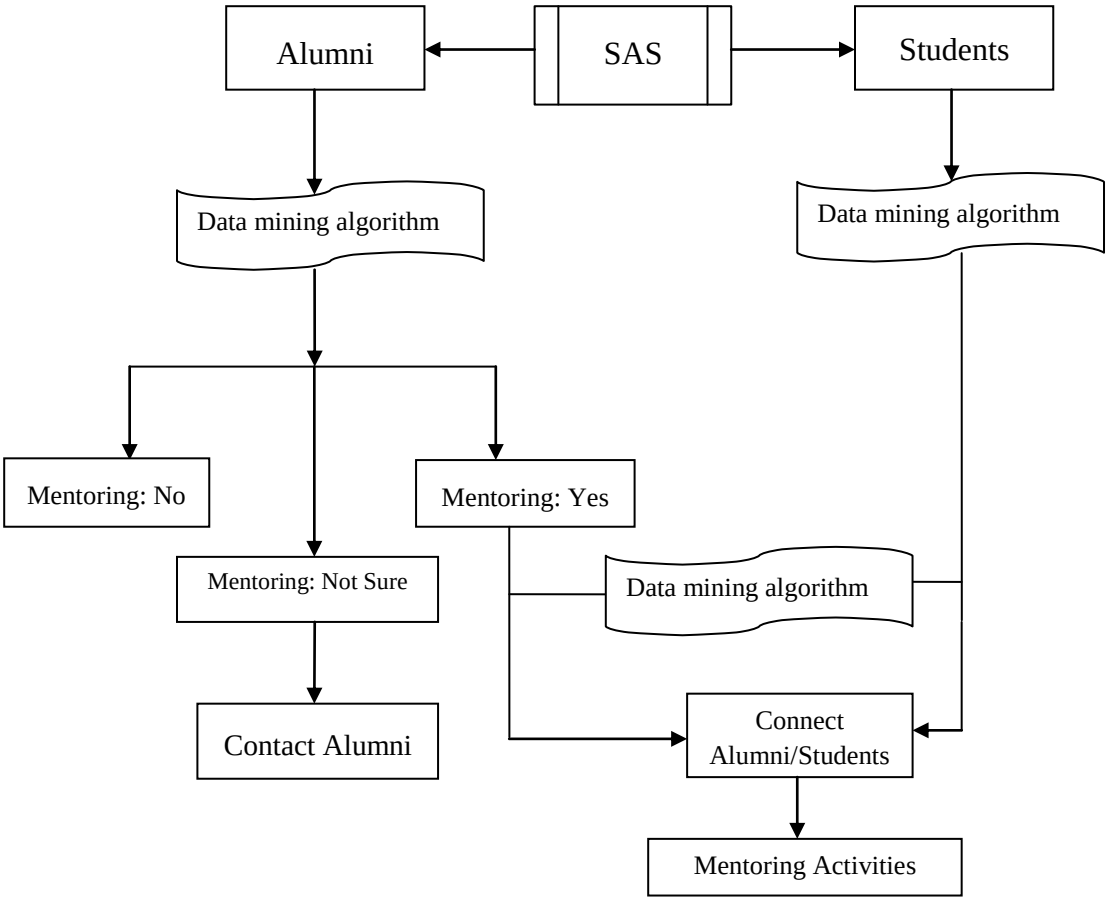

Figure 11: Schematic for mentoring process

students need to find jobs, they can concentrate on the companies and firms who currently host their senior alumnus from the department or the university. The alumni can help the current students by giving information of the openings in the company, about the work culture and the key persons whom they can approach, without crossing the guidelines of the company. All such interactions can be done on a one-by-one approach where only a single student contacting the alumni is benefited. Alternatively, the alumni could try to convince their employers to visit the universities where they graduated when in need of new recruits. The process is shown in Figure 12. In SAS all the alumni in the

system can be divided in accordance to the employer they are working for, their title/position and the university can contact the alumni and ask them to encourage their employers to schedule a visit for recruitments. This will be great help for the current students as it would reduce the time to find a job.

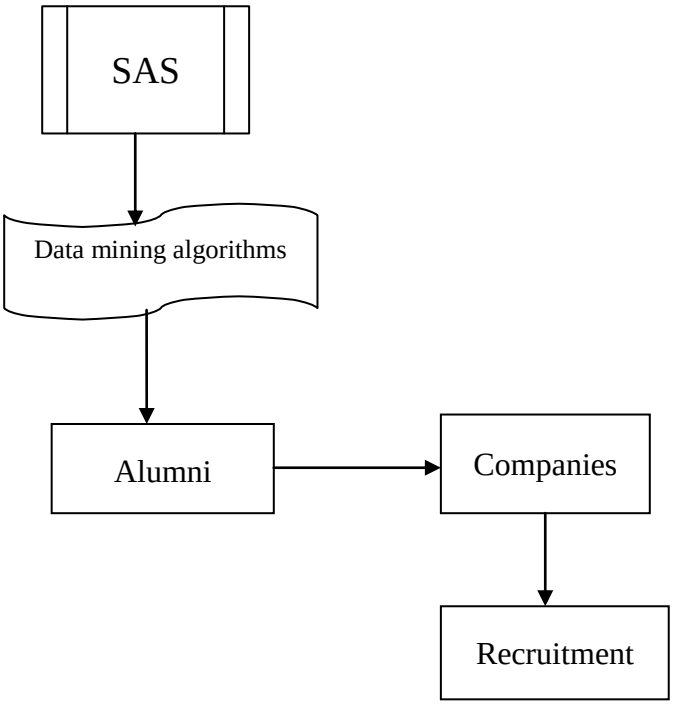

Figure 12: Schematic for recruitments

As shown in Figure 12, alumni in SAS can be mined to group the alumni along various attributes like the company they are working for, their designations, and years they have been working for a specific company, etc. Depending on all the factors, the appropriate alumni can be contacted who can be able to convince their employer to visit the institution for recruitment. This is even advantageous for the recruiting companies as they can get more potential employees from recruiting in a single place and, before

coming for the recruitment, the companies will have an idea of the caliber of the students from the alumni working for them.

4.2.5 Alumni Career Paths

Smart Alumni System contains alumni information ranging from the year of joining the university, year of graduation, their initial occupation to their current work profile.

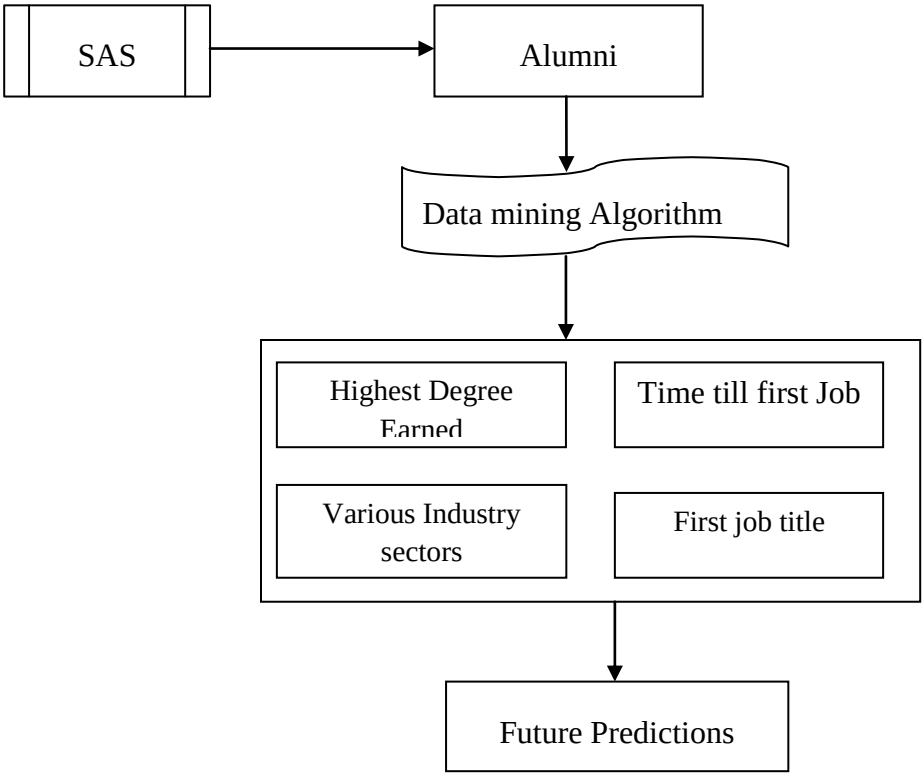

Figure 13: Schematic for process of tracking alumni career paths

For the better development of the university, SAS should be able to obtain the career path of alumni which can be helpful in different ways. For example, the alumni graduating in

electrical engineering may choose to become a software professional, while some might choose to become teachers and educators. The graduating students can make use of these pre-opted career choices, and try to focus or if necessary change their job search strategy. If the students with the undergrad in a particular university are getting a master degree from another university, they can try to analyze the reason for this step. It can also help the universities to put the future career of the current students in perspective motivate them through predictions based on alumni who have already been through the process before.

4.2.6 Helping Guests

In SAS there is a provision for the guests to be members of the system. The guests can be prospective students, their parents, recruiters etc. The type of interaction they have within the system or with the system will make an impression on them about the university. It is always advantageous for the universities to have more potential students enroll with it as it would increase the reputation and revenue of the university. Guest users in SAS can interact with the faculty and current students in the system. So when a guest registers with SAS, depending on the purpose of the visit to the system, they will be guided to appropriate people or resources. For example, if a parent registers with the system for knowing more about the university and a specific department, SAS can send them information about the corresponding faculty members who can be of more helpful. Likewise, when prospective students register with SAS, they can have access to faculty and current students to whom they can ask their questions about program, courses, university living and activities. Data mining also allows us to keep track of the number of

guests who become associated with the university as well as to run predictions depending on the various data available about the guest which helps the university to concentrate in aspects relevant to attracting a multitude of visitors. In the design below, data mining can determine the best faculty and student who match the requirement of the guests and the guest can contact them for getting more information.

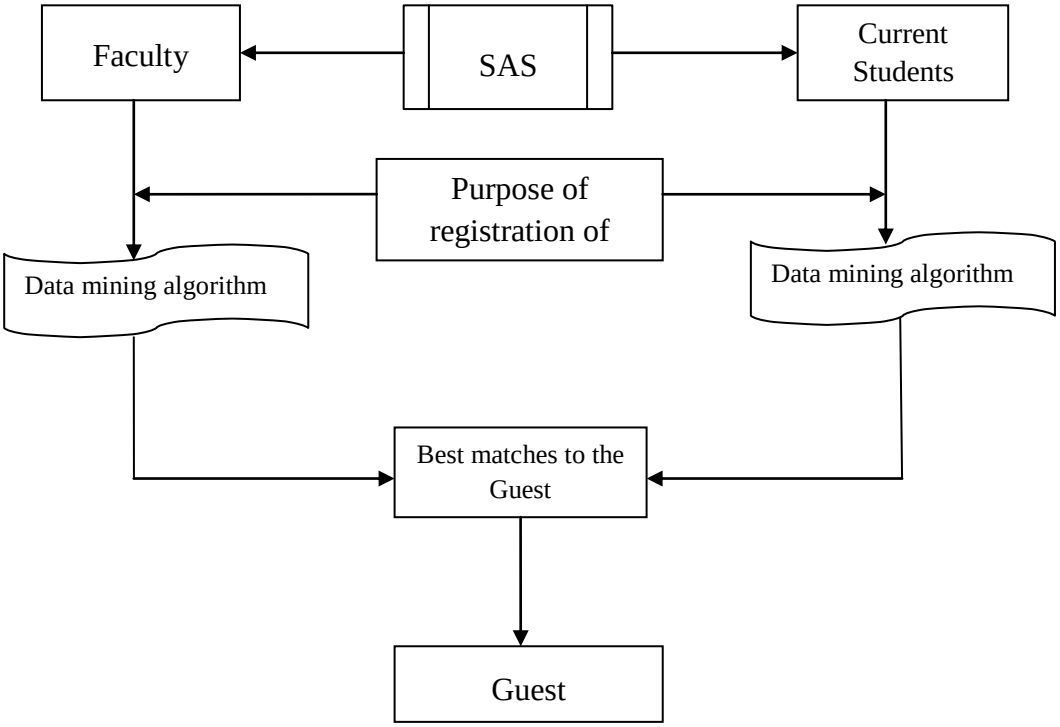

Figure 14: Schematic for process of helping guests

4.3 Algorithms

The data mining algorithm is the mechanism that creates a data mining model. To create a model, an algorithm first analyzes a set of data and looks for specific patterns and trends. The algorithm uses the results of this analysis to define the parameters of the

mining model. These parameters are then applied across the entire dataset to extract actionable patterns and detailed statistics [15].

The mining model that an algorithm creates can take various forms, including:

- A set of rules that describe how products are grouped together in a transaction.
- A decision tree that predicts whether a particular customer will buy a product.
- A mathematical model that forecasts sales.
- A set of clusters that describe how the cases in a dataset are related.

Choosing the best algorithm to use for a specific business task can be a challenge. While you can use different algorithms to perform the same business task, each algorithm produces a different result, and some algorithms can produce more than one type of result. Data mining functions used in the Smart Alumni System are: prediction, clustering, classification and association.

4.3.1 Predictions

In the prediction function, the goal is to develop a model which can infer a single aspect of the data (predicted variable) from some combination of other aspects of the data (predictor variables). Prediction requires having labels for the output variable for a limited dataset, where a label represents some trusted "ground truth" information about the output variable's value in specific cases. In some cases, however, it is important to consider the degree to which these labels may in fact be approximate, or incompletely reliable. Prediction has two key uses within educational data mining. Firstly, prediction

methods can be used to study what features of a model are important for prediction, giving information about the underlying construct. This is a common approach in programs of research that attempt to predict student educational outcomes [16] without predicting intermediate or mediating factors first. In a second type of usage, prediction methods are used in order to predict what the output value would be in contexts where it is not desirable to directly obtain a label for that construct [17].

4.3.2 Clustering

In clustering function, the goal is to find data points that naturally group together, splitting the full dataset into a set of clusters. Clustering is particularly useful in cases where the most common categories within the dataset are not known in advance. If a set of clusters is optimal, within a category, each data point will in general be more similar to the other data points in that cluster than data points in other clusters. Clusters can be created at several different possible grain-sizes: for example, schools could be clustered together (to investigate similarities and differences between schools), students could be clustered together (to investigate similarities and differences between students), or student actions could be clustered together (to investigate patterns of behavior) [18, 19]. Clustering algorithms can either start with no prior hypotheses about clusters in the data (such as, the k-means algorithm with randomized restart), or start from a specific hypothesis, possibly generated in prior research with a different dataset (using the expectation maximization algorithm to iterate towards a cluster hypothesis for the new dataset). A clustering algorithm can postulate that each data point must belong to exactly one cluster (such as, in the k-means algorithm), or can postulate that some points may

belong to more than one cluster or to no clusters (such as, in the case of Gaussian mixture models)

The goodness of a set of clusters is usually assessed with reference to how well the set of clusters fits the data, relative to how much fit might be expected solely by chance given the number of clusters, using statistical metrics such as, the Bayesian information criterion.

4.3.3 Classification

Classification is a major data mining operation that takes a given attribute and predicts the values of the class using the other available attributes. These apply to categorical outputs for an attribute which takes on two or more discrete values, also known as a symbolic attribute.

The most frequently used algorithm in classification are decision trees. A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. Another use of decision trees is as a descriptive means for calculating conditional probabilities. When the decisions or consequences are modeled by computational verb, then we call the decision tree a computational verb decision tree [20]. The decision trees are considered when instances describable by attribute - value pairs; target function is discrete valued; disjunctive hypothesis may be required; possibly noisy training data. First, select an attribute to place at the root node and make one branch for each possible

value. This splits up the example set into subsets, one for every value of the attribute. Now the process can be repeated recursively for each branch, using only those instances that actually reach the branch. If at any time all instances at a node have the same classification, stop developing that part of the tree. The only thing left to decide is how to determine which attribute to split on.

Any leaf with only one class not have to be split further and the recursive process down that branch will terminate. Because small trees are needed, it is better to happen as soon as possible. If the purity of each node is to be measured, the attribute that produces the purest daughter nodes should be chosen. The measure of purity that being used is called the *information* and is measured in units called *bits*. Associated with a node of the tree, it represents the expected amount of information that would be needed to specify how a new instance should be classified [21].

Table 3: Steps for basic classification algorithm

Main loop:
1. A the “best" decision attribute for next node
2. Assign A as decision attribute for node
3. For each value of A, create new descendant of node
4. Sort training examples to leaf nodes.
5. If training examples perfectly classify, Then STOP, Else iterate over new leaf nodes

4.3.4 Association

Associations are used to find correlations between different attributes in a dataset. Association rules are like classification rules. They can be found in the same way, by executing a divide-and-conquer rule-induction procedure for each possible expression that could occur on the right-hand side of the rule. But not only might any attribute occur on the right-hand side with any possible value; a single association rule often predicts the value of more than one attribute. To find such rules, rule-inductions procedures have to be executed once for every possible combination of attributes, with every possible combination of values, on the right-hand side. That would result in an enormous number of association rules, which would then have to be pruned down on the basis of their coverage (the number of instances that they predict correctly) and their accuracy (the same number expressed as a proportion of the number of instances to which the rule applies). This approach is quite infeasible. Typically, association rules with high coverage are considered. The distinction between the left- and right-hand sides of a rule and seek combinations of attribute–value pairs that have a pre-specified minimum coverage. These are called item sets: an attribute–value pair is an item. This is mostly and conveniently used for market basket analysis, in which the items are articles in customers shopping cart and the supermarket manager is looking for associations among these purchases.

4.4 Data mining tools

The main aspect of data mining is the usage of algorithms. Data mining is about specifying patterns, and algorithms to find them. The usual procedure analyzes different algorithms, and the best algorithm which suits the specific function is used. This section

gives an overview of the different tools that can be used for data mining. There are different tools which are developed by various organizations which have these algorithms built-in. When the dataset is given to the tools the user can select an algorithm and the tools to develop the models. Some of the good tools available for free are *Weka, Orange, Rapidminer, and Statistica* [22-24].

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. *Weka* contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes [23]. This tool is named after *Weka* which is an endemic bird of New Zealand. *Orange* is a component-based data mining and machine learning software suite, featuring friendly yet powerful and flexible visual programming front-end for explorative data analysis and visualization, and Python bindings and libraries for scripting. It includes comprehensive set of components for data preprocessing, feature scoring and filtering, modeling, model evaluation, and exploration techniques. *Orange* was developed at the Bioinformatics Laboratory of the Faculty of Computer and Information Science, University of Ljubljana, Slovenia [24]. *RapidMiner*, formerly YALE (Yet Another Learning Environment), is an environment for machine learning, data mining, text mining, predictive analytics, and business analytics. It is used for research, education, training, rapid prototyping, application development, and industrial applications. *RapidMiner* provides data mining and machine learning procedures including: data loading and transformation (ETL), data preprocessing and

visualization, modeling, evaluation, and deployment [25]. *Statistica* is a statistics and analytics software package developed by StatSoft. *Statistica* provides data analysis, data management, data mining, and data visualization procedures.

5. DESIGN FOR SMART ALUMNI SYSTEM

In this chapter, the design and structure of Smart Alumni System is described in detail.

5.1 Structure of SAS

The Smart Alumni System developed as a part of this thesis is based on a three-tier architecture comprising of an interface, middle ware and databases. Figure 15, shows the three-tiered architecture of SAS. The user interface layer permits users to see and access the system. The middle layer contains code used to access the database and enforce system access control. Middleware also includes the programming logic for data mining. The database layer manages all system data — information entered by the user, data obtained from external sources and data produced from various data mining tasks.

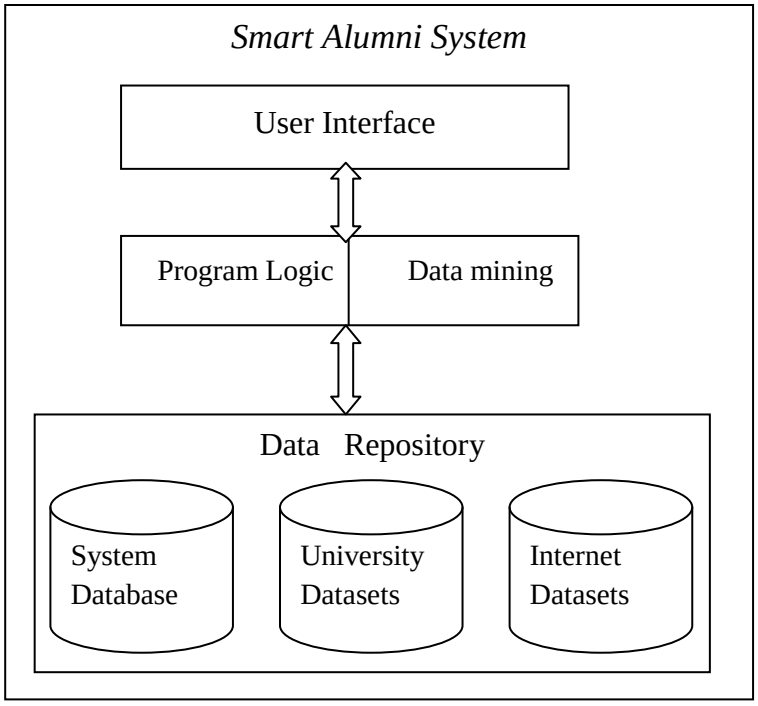

Figure 15: Structure of Smart Alumni System

5.2 Interfaces

This section gives an overview of the interfaces in SAS. As mentioned above, the SAS is a web based system. The users need to login to the system in order to access it. Below are a few model interfaces of the system:

Homepage and Login. Homepage is system's main screen that has provision for access into the system; typically through a login tab. Users registering with SAS may be required to use their university email to preserve the exclusivity of the system. Guest users, who by definition have limited access, may be allowed to login using alternate emails. Such an identification of user group, based on their email in turn helps to configure the access definitions of users in SAS and streamline the permissions.

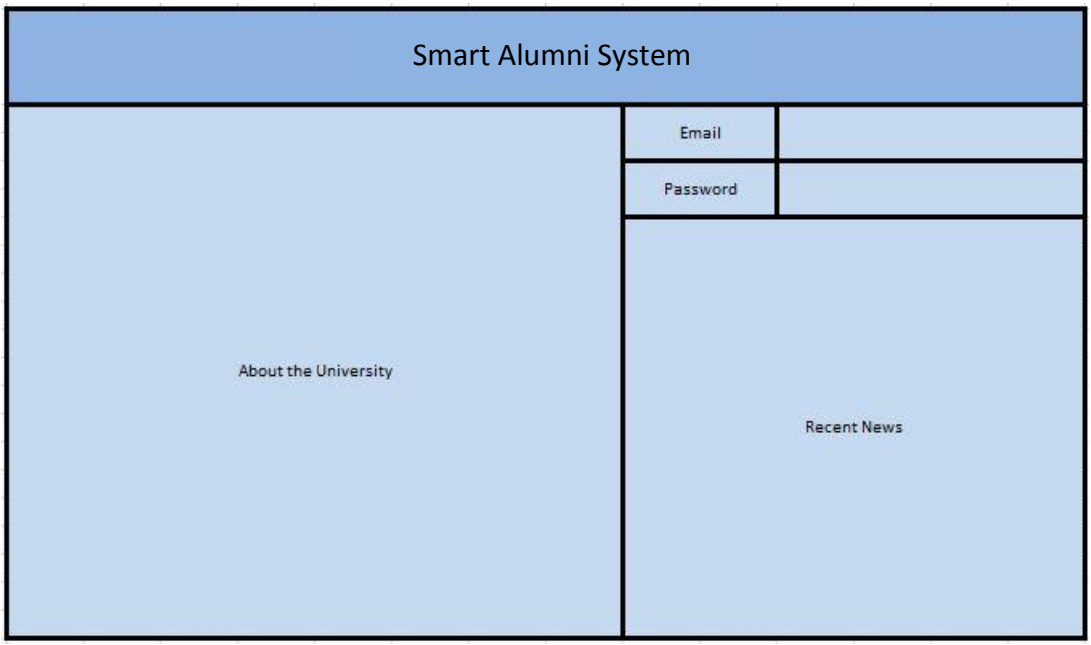

Figure 16: Typical interface of home page for login authentication

Registration page. A registration process is mandated for every user of the system. Figure 17, shows typical data a user must provide during registration. Based on the details entered at registration, a member is automatically classified into a user category and his privileges are set to corresponding values. An additional layer of authentication can be incorporated by allowing a SAS administrator to examine the registration data. Members might also be asked to provide answers to a few security questions aimed at identity verification.

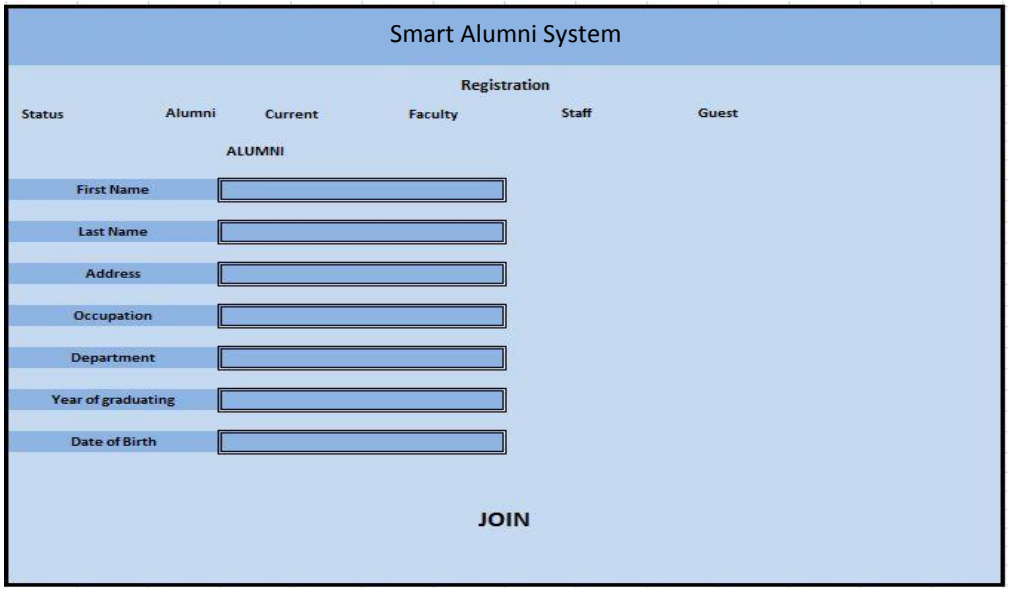

Figure 17 : Web interface for registration

Profile page. The alumni, current students and faculty have a dedicated profile page wherein they can enter information they would like to share with the other members in system. Based on its relevance the information may be sectioned as: general information, social information, contact information, and professional information. The member is

provided with control features which allow him to specify the amount and type of information one would like to share.

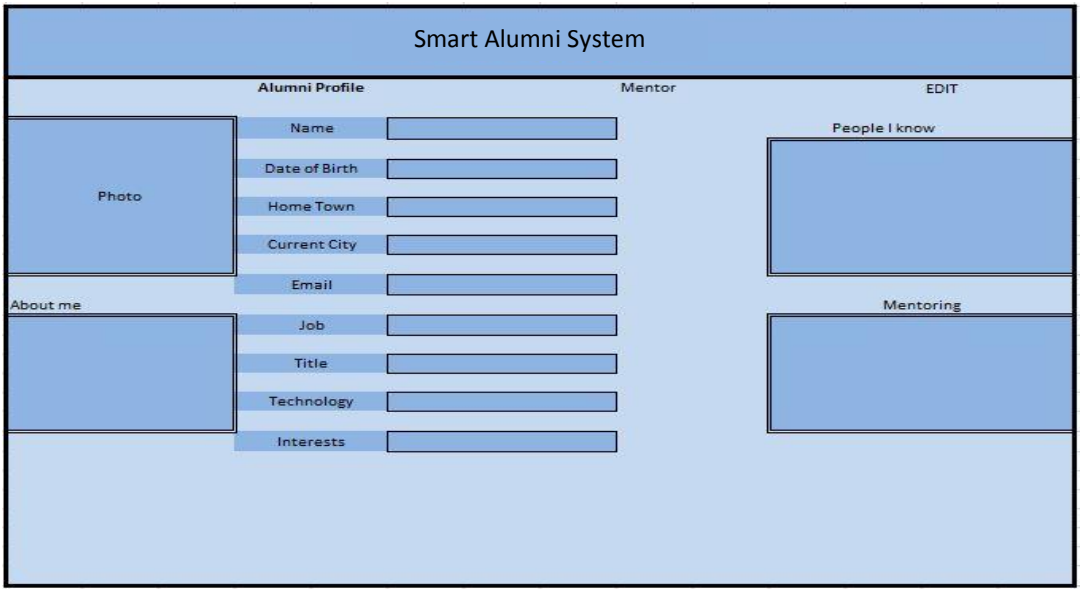

Figure 18: Model interface of profile page in SAS

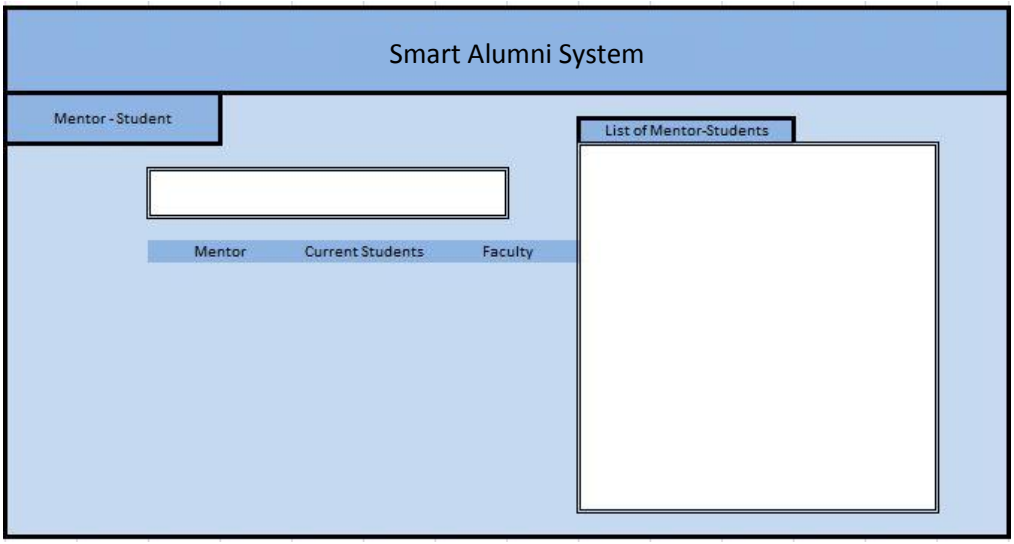

Figure 19: Model interface of search page in SAS

Search page. Users of SAS are provided with a search feature that allows them to look for their acquaintances in the SAS members. Search can be based on various fields: occupation, city, major and year of graduation. A typical search would yield a list of member profiles which match the specified search fields. Since, SAS has features which are akin to those available on social networking websites, the users can not only search to find friends but also form groups, and have dedicated blogs for discussions on common interest.

5.3 Data Repository

The database layer essentially holds all the data of SAS. It comprises of — user data; data obtained from university records; data obtained from thirty parties like social networking websites.

User data. Information entered by the user during registration or later is one of the key datasets of SAS. It can include all personal, educational, professional and miscellaneous details about the member.

University datasets. Most universities tend to conduct surveys from graduating class on wide range of related to educational standards and expectations at the university. The volumes of data collected from such surveys can be analyzed to help the university administration with key decisions related to improving the education experience at their institution. The data repository can be used to store both raw data from university surveys and processed data obtained after analysis.

Internet datasets. Many social networking websites collect and analyze data. Integration of these datasets with SAS data would expand the research scope of SAS and benefit all the stakeholder of it. For example, a SAS member might be able to search and invite prospective SAS members among people who are currently registered with social networks.

Analysis of data obtained from university or social network repositories can be yields additional business intelligence. An SQL server would handle such databases and any query pertinent to it. Data saved in database is tabular and can be retrieved into an Excel sheet or '.csv' file format.

5.4 Use Case Diagrams

The essential interactions in SAS can be described with the help of use case diagrams. Use case diagrams provide a graphical overview of the functionality provided by SAS, in terms of a description of the steps or actions between a user and the software system with which the user interacts. The following sub sections describe the use case diagrams for interaction requirements between the faculty, alumni and current students.

5.4.1 Use Case Diagram for Alumni

Alumni play a primary role in SAS, as most features of SAS are concentrated on how alumni relations with the university can be enhanced. As shown in Figure 20, SAS provides many functions to registered alumni aimed at supporting their activities of interest. For example, interested alumni can act as mentors for current students and guide them with career choices; they can also help organize recruiting events. Alumni can also contribute to curriculum development by participating in discussion boards and groups

aimed at curriculum improvements. SAS provides a secure system which can help determine the alumni who can donate or pledge money for development of the university or individual departments.

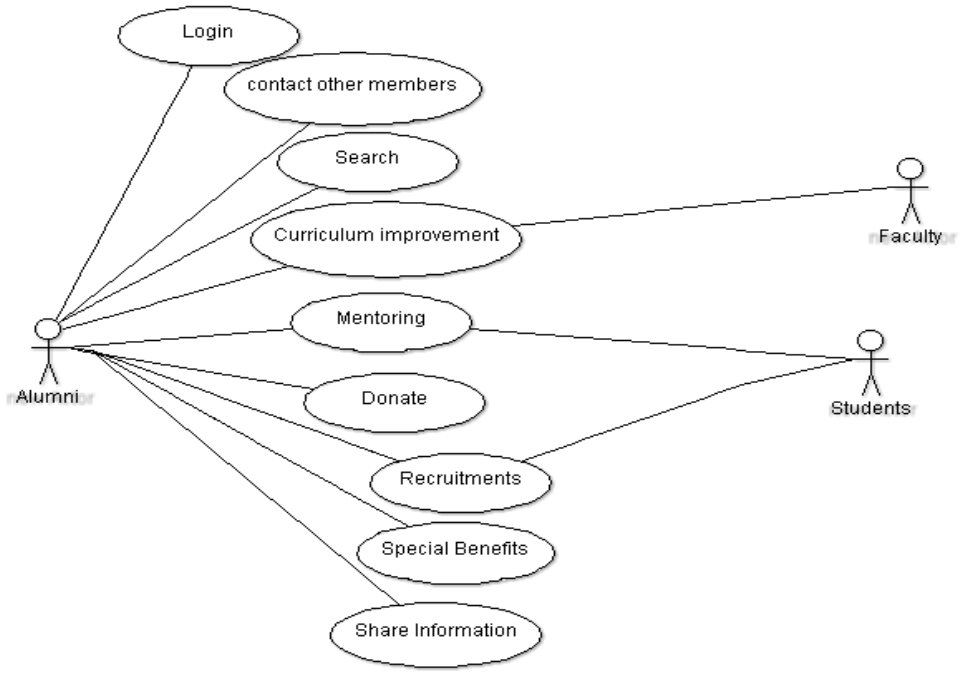

Figure 20: Use case diagram for alumni members

5.4.2 Use Case Diagram for Students

Unlike traditional alumni sites, current students of the university can also be a part of the SAS. Students are allowed to participate in SAS keeping in mind the possibilities of forging links between current students and alumni, a step which could provide current students with mentors or connections instrumental in finding a job. Current students are allowed to interact with faculty members and participate in social forums. As such, functions which allow access of information between current students and the alumni

body are required. Figure 21, shows use case diagram specific for student member needs as a part of SAS.

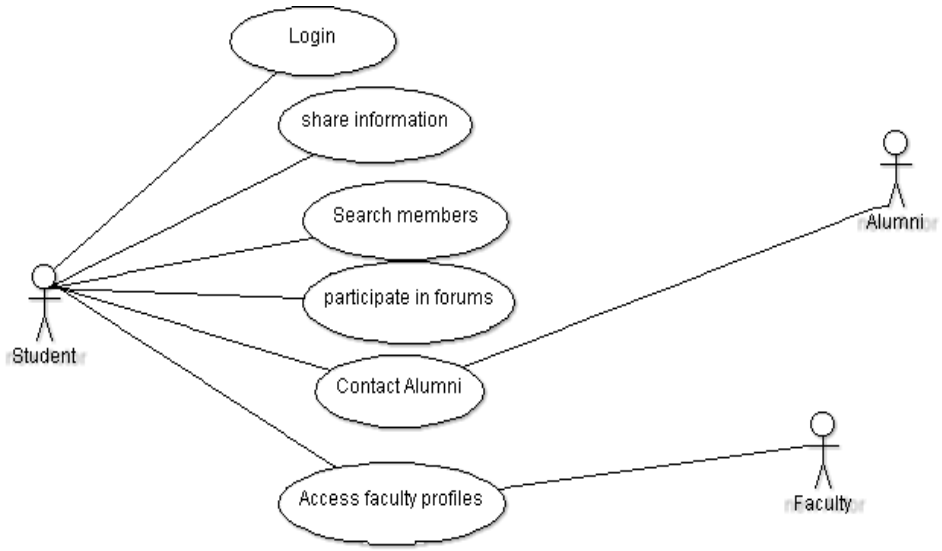

Figure 21: Use case diagram for student members

5.4.3 Use Case Diagram for Faculty

SAS provides faculty with an environment to maintain relations with the current students and alumni who are registered with SAS. Since SAS is not a class room environment and involves one-to-one communication there is a better chance that students share their concerns and give feedback to faculty. Based on such interactions, the faculty can organize study groups, have discussion forums and also attract students who are interested in the faculty's research. The use case diagrams for these faculty specific interactions are shown in Figure 22.

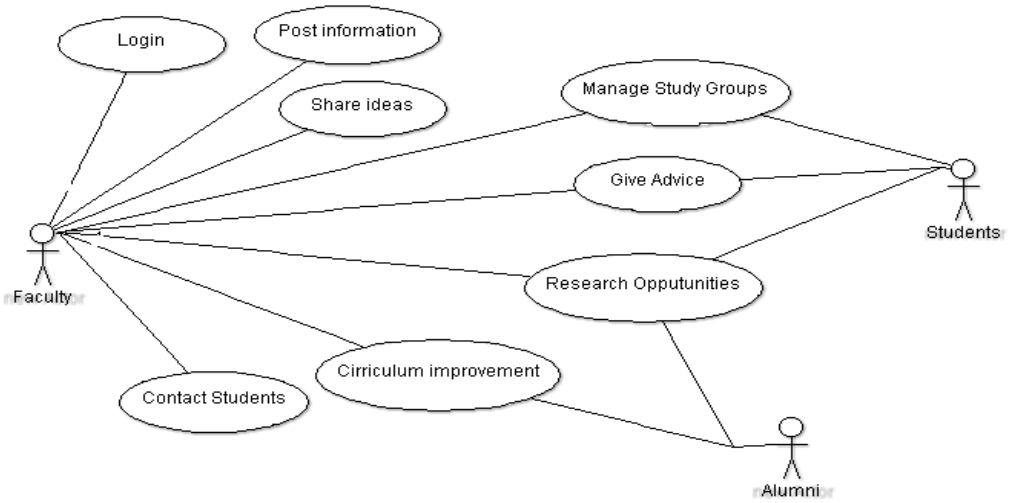

Figure 22: Use case diagram for faculty members

5.5 Case Study – CIS Alumni System

As a case study of the alumni system of the CIS (Computer and Information Sciences) Department at Florida Agricultural and Mechanical University is considered. The CIS Department began as an academic program in 1967 and has graduated over 1200 students since its inception. At the 40th anniversary celebration of CIS during the 2007-08 academic years, a dedicated CIS Alumni System (CAS) website was developed for CIS alumni. Its homepage can be found at http://www.cis.famu.edu/~alumni. Although CAS is an open system, it does not support many activities. Also, the information collected from stakeholders (e.g., alumni surveys) is updated manually. Much of the collected information is openly accessible. The structure of the information (i.e., database) is not sufficiently clear. Many improvements to CAS can be envisioned. The CAS system does not have the salient features of SAS, e.g., current student participation, faculty, visitor and other interactions. Most alumni systems have a similar structure and are typically

restricted to alumni. Developing a Smart Alumni System will provide better interaction mechanisms all stakeholders, which is expected to encourage members to use the system more frequently.

5.6 Uniqueness of SAS

Currently, there are many different types of the alumni systems in use by different universities. One common feature observed in all the traditional alumni systems is that the membership is only limit to alumni. SAS developed in this thesis expands the membership landscape and offers wide-range of interaction functionalities, which can further studied to make the system efficient and enhance the user experience. The uniqueness of SAS can be identified as under:

Stakeholders. SAS has a expanded list of stakeholders, these include: current students, alumni, faculty and staff. It aims to be a one stop communication point for all the involved stakeholders.

Exclusive Network. SAS is an exclusive network of people who are affiliated with the university either in the present or the past. Although guest members like visitors and company recruiters can get into the system their access is very limited and SAS users need not be concerned about strangers prying on their information.

Social Networking. Unlike the traditional alumni systems, SAS provides the members user profiles which can be updated by the user, and its content can shared, a feature much like the ones provided by social networking websites. Such features are likely to make SAS very attractive, given the popularity of social networks and the ease

of user getting adapted to it. Also, features such as, search and add friends also facilitate expansion of user base.

Privacy settings. Since SAS is an exclusive network, only authenticated students, alumni and university personnel are members of it. Any stakeholder who would like to be a member of SAS needs to register, and all registration requests are screened by a SAS administrator who in turn checks the university database and concerned departments to make sure that the individual stakeholder is indeed associated with the university. In addition, all the data collected is securely handled by university only for its specified research objectives, thereby eliminating risks involved with sharing of information and advertising spam.

Data mining feature. The use of data mining techniques in the SAS is one of its most distinguished features. The application of data mining algorithms to understand patterns of interest is what makes the alumni system to be 'smart'. Patterns observed in the data would benefit the members of SAS and contribute to overall development of the university. A few examples on use of data mining for such applications were discussed earlier in chapter 4.

6. IMPLEMENTATION OF SMART ALUMNI SYSTEM

This chapter elaborates on the implementation details of SAS, strategies used to develop the system and application of data mining to key questions of stakeholders interest.

6.1 User Interfaces

The current implementation of SAS is built for three stakeholders: current students, alumni and faculty. Users in the system can communicate with each other and access member information based on their classification. The key user interfaces available in the system are described below.

6.1.1 Login Page

The SAS login page serves two purposes. First of all, it enables a user to login to the system. Secondly, it provides a preview of the university website. If desired, the user can navigate though the university website to know more information about the programs, departments and university related news. Such feature is more useful to guests who would like to know more about the university.

All first time users need to complete a 'registration form' obtained via the 'New User' tab. The registration window collects mandatory information like first name, last name, email, affiliation status (student, alumni, and faculty) and date of birth; which is then checked with university records to establish user authenticity. A screenshot of the login screen of FAMU-SAS is shown in Figure 23.

Figure 23: Screenshot of login page in SAS

Once the registration process is successful, the user can use his email and password to log into the system using the login portal. In case the user forgets his password he is allowed to recover it using the 'Forget Password' link.

6.1.2 Profile page

Profile page is the portal which contains user information which the user would like to be displayed on his profile and provides control options to specify which content can be shared with other members of SAS. User information is organized into three different categories and displayed under separate tabs: Profile Details, Personal Details and Work Details.

Profile Details: The 'Profile' tab contains general information entered by the user during registration. Fields in this tab inherit basic details like first name, last name and affiliation status entered during the registration process, and provides fields for entering a brief description about the user.

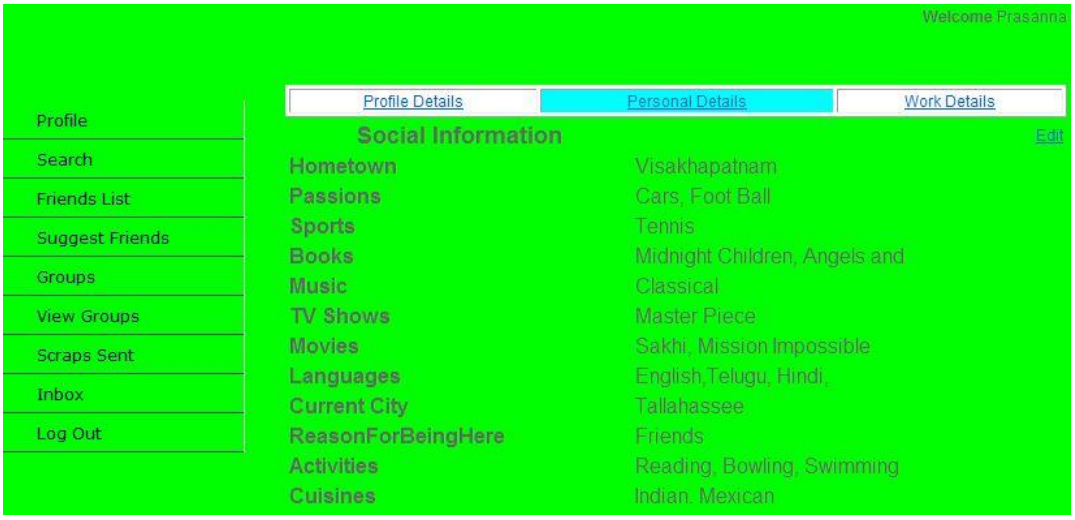

Figure 24: Screenshot of profile Page in SAS

Personal Details: The 'Personal Details' section in the profile page contains information of the user like Hometown, Passions, Sporting interests and favorite TV shows. Most of the fields in this tab are not mandatory and the user can edit them at his or her convenience. Users can also use this tab to customize personal information they want to share with other members in the system. Data collected in this tab forms an important part of user data since it is later used to make connections between the members of the system.

Work Details: The 'Work Details' tab in the profile page contains information related to the user's education, employer and other career related details. Since, the main

motto of SAS is to build a career based networking for its members; this tab is where most of the important user information is collected. This tab is sub-divided into two parts: educational details and career details. 'Educational Details' contains a summary of the educational background of the member: schools attended, the highest degree earned, the year graduated, majors and minors received. The 'Career Details' tab contains information about the jobs undertaken, brief description of work details and other relevant information. It should be noted that both 'Work Details' and 'Personal Details' are optional.

6.1.3 Friends page

The 'Friends' page of a user lists all the members of SAS who are friends with the user, and provides links to their individual profiles. It also shows links available to initiate a communication with a friend, shown in Figure 25.

Figure 25: Snapshot of friends list tab in SAS

SAS provides a ‘Suggest Friend’ feature popularized by social networking media. It allows the user to search for his friends and acquaintances and to add them to his friends list within a few clicks. This tab is linked with data mining routines to recommend likely acquaintances and friends of the user based on the information provided in user profiles. The ‘Suggest Friend’ tab provides two unique options for each user type. The options for students are — friends, company; for alumni are — friends, mentoring; for faculty are — funding, curriculum development. The processes used to make suggestions for each type of user are explained in section 6.3.

6.1.4 Search Tab

The search tab enables the users to search for members within the system. A search can be based on any of six distinct fields which are available to the user. The fields on which the search can be based are: username, first name, last name, high school, home town and occupation, shown in Figure 26.

Figure 26: Snapshot of the search page in SAS

6.1.5 Groups

Another unique feature provided by SAS, when compared to traditional alumni systems, is the ability to form virtual groups. To form a group a user would navigate to the groups tab and enter a title for the group, brief description and its type (optional). The group type can range from educational, social, sports, outdoor activities to any other activity. Once created the group a user can be share it with his friends in SAS. The members of the group can also see the particular group listed under the 'View Groups' tab.

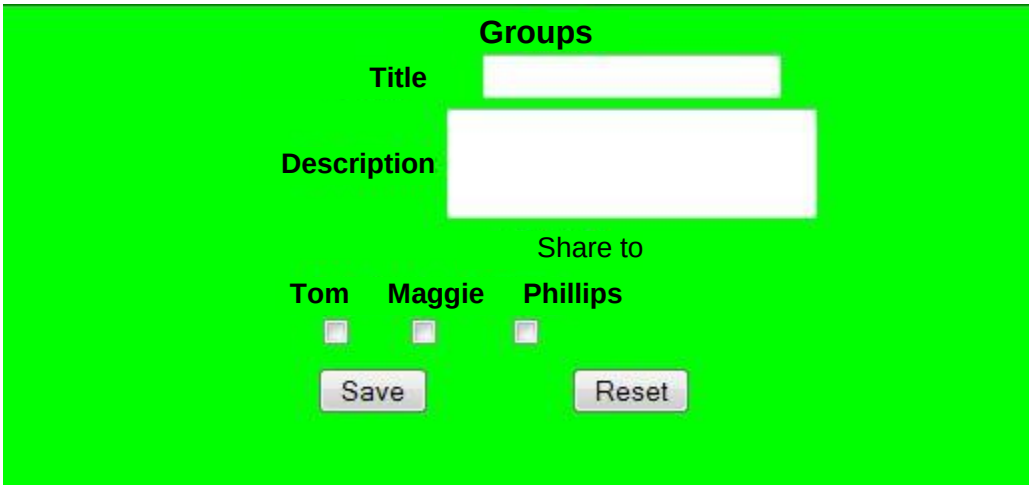

Figure 27: Snapshot of groups tab in SAS

6.1.6 Messages

Two tabs exist in SAS which are related to message communication. A tab titled 'Scraps Sent' displays all the messages sent by the user. The other tab titled 'Inbox' displays the messages received by the user. To send or receive messages, the sender and the receiver need to be friends with each other.

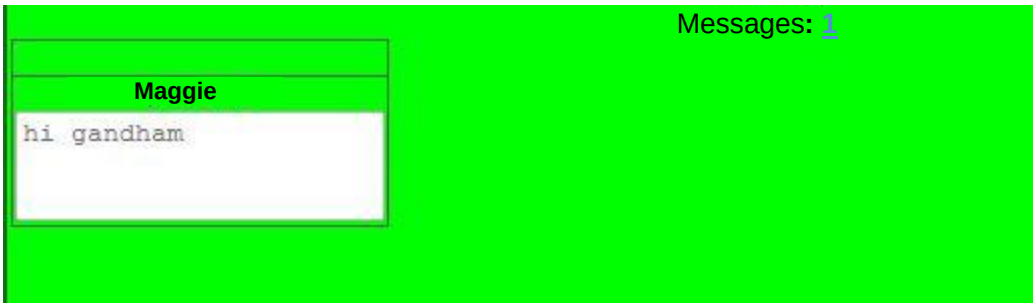

Figure 28: Screenshot of inbox where members can receive messages

6.2 Use of Data mining

Data mining algorithms are used in SAS to analyze volumes of data collected from all the users, and to assist the user through suggestions in various aspects which depend on the user status. Features specific to the stakeholders — current students, alumni and faculty, are discussed below.

6.2.1 Students – Friend suggestions

Current students are given friend suggestions based on the data entered by the student and its match with other student members of SAS. All the information is analyzed using a ranking algorithm which investigates the system data for patterns, from which a model is developed. When new student member data are available to the system, the model is used to find out suggestive outcomes. Using the ranking algorithm, a model with ranks for all fields is obtained, which helps SAS in devising more accurate suggestions. While preparing suggestions, the fields with highest rankings are considered as priority.

The datasets used for data mining can be collected from various sources like the information entered by the user during registration, data from the universities and data from social networking sites. In the present study, only data entered by the user are

considered (extensions to include other forms of data will be undertaken in the future). It should be noted that a data mining algorithm produces results that are different from a simple search. A simple search scans through the whole data using just the keyword, and an output is obtained only if the keyword matches. On the contrary, data mining algorithms linked to the ‘Suggest Friends’ tab scans though the whole data and instead of depending on a single keyword, it uses various fields to make suggestion for the user. The fields considered for giving friends suggestions for the students are hometown, department, age and major, in that particular order.

6.2.2 Students – Career suggestions

Current students can make better use of SAS to seek career guidance. If a current student wants to find alumni who work with a specific employer, he can go through the profiles of all alumni members in the system and try to find who can be contacted. SAS makes this process convenient for current students by providing them with good suggestions on alumni contacts. The student can select the ‘Company’ link in the ‘Suggest Friends’ tab to get corresponding members who are working for a certain company of interest to the student. Such suggestions are obtained using ranking algorithm which goes through all the alumni data in SAS and finds out high ranked fields in the model. The fields on which the suggestions are based for a company selection are hometown, department, current city and career interests, in that order or relevance.

6.2.3 Alumni – Friend suggestions

SAS is an ideal place for alumni to get in touch with their old friends. In this case, the ranking algorithm applied to find friends for current students is adopted with an

exception that it uses only alumni data when giving suggestions to the alumni. The fields considered to offer suggestions to the alumni are their department, graduation year, major and home town; in that order of priority. Owing to the difference in data mining models, and the data itself, the ranking of the fields for alumni is different from that of current students.

6.2.4 Alumni – Mentoring

Mentoring is another main objective of SAS. Alumni who are willing to help current students can volunteer to do so. An alumnus who assumes the role of a mentor can help current students on technical or non-technical matters, and can guide them in their career paths. For finding mentoring related data mining, the association algorithm (described in section 4.3) is used. The current students who need mentors have to make a request for a mentor in the SAS. The alumni interested in mentoring the students need to express their willingness, at registration or at a later point. Upon availability of data from both current students and alumni, the association algorithm tries to pair an alumnus with a current student based on their interests and background.

With regards to mentoring, all the fields in the profile are ranked according to match between mentors and the students. The fields considered include: their department, interests, hometown, career interests, and employer the alumnus is associated with, in that particular order of priority.

6.2.5 Faculty – Donations

Faculty and staff members associated with the university also benefit from data mining features, since it helps them connect with former students as well as current

students. For example, SAS facilitates connections between faculty and former students aimed at securing funds, grants which contribute to development and improvement of the university, department or research program.

SAS has a dedicated page to receive alumni donations. Donation information of the alumni members is available to faculty and administrative members who use it to obtain suggestions from SAS in the form of a list of alumni who are more likely to donate. In implementation, this feature is incorporated into the faculty profile. SAS makes use of ranking to predict alumni tendencies to make a donation. Donation data in the system is analyzed and a model is developed to find out which alumni have already contributed to the university in the past. The model is then used to identify new alumni who are likely to donate. Alumni who graduated long back are more likely to be well placed and financially solvent to make a donation. Hence, alumni users in that group will be a better choice to contact. The order of ranks for the fields with respect to fundraising are year graduated, company, home town, department and position.

6.2.6 Faculty - Curriculum Development

There is a need for universities to stay abreast with the new emerging technologies and industrial practices. As such, it is required to modernize the curriculum used to train students and equip students with better skills for the job market. While considering upgrades to curriculum the best people to contact are the alumni who are already working in various technological sectors of the industry. Using data mining techniques, the faculty can be teamed up with the right set of alumni who can give valuable inputs for curriculum development. One way to accomplish this objective would be to shortlist the

alumni who work in trades related to a specific course (which needs improvement) and seek their advice on updating the course structure or its contents. In this case, the alumni who might be present in the same place or nearby might be more participating due to their geographic convenience in getting involved. So for curriculum development the fields considered are department, current city, career interests, company and position; in that order of priority.

7. CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

A Smart Alumni System is an online networking system where the membership is not confined just to the alumni but is extended to the current students, faculty and other members associated with the university. In this thesis, a Smart Alumni System (or SAS) that is far more outreaching then traditional alumni systems has been designed and implemented. This project has demonstrated that traditional systems can be re-engineered with aspects of social networking keeping in view the potential benefits and success of social networking websites in capturing user attention and involvement. Inspired by user functionalities in social networking media, a framework for SAS is outlined. Further, it has been shown that data mining principles can be applied to user data obtained from alumni data to enhance the functionality of the SAS.

The proposed Smart Alumni System has been implemented using ASP.net with C#. The implemented system provides a platform for communication and active interaction between alumni as well as current students, faculty, and other personnel affiliated with the university. Each stakeholder is classified into user sub-groups of a specific kind whose permissions and functionalities are pre-defined. Users of different kind are allowed to interact among themselves and also with users from different sub-groups in order to achieve common objectives or mutually beneficial tasks. It has been shown that proposed Smart Alumni System can benefit from data mining principles in order to provide an enriched experience and interactive functionalities to users of the system.

7.2 Future Work

The following are the possible directions for future research:

- The current implantation of SAS can be expanded to include data obtained from university records and third party social networking sites.
- The data mining principles used for SAS can be applied in conjunction with fuzzy logic algorithms to yield a more probabilistic model based on available datasets. Such a step would take into account more possibilities of connections between users of the system.
- Another area of research is addition of semantic web features to allow user data to be shared and reused across different applications and system boundaries. For example, if the user updates current city in a social networking site the semantic web features would enable an automatic update of the city in SAS, vice versa.
- Finally, SAS can be made more readily accessible on mobile technologies like cell phones and handheld devices by developing exclusive applets for this purpose.

REFERENCES

1. Schomburg, H., Teichler, U., "Increasing Potentials of Alumni Research for Curriculum Reforms: Some Experiences from a German Research Institute", *New Directions for Institutional Research*, Vol. 2005(126), 2005, 31-48.

2. Trinklein, D., Wells, J., "Use of Alumni Survey in Curriculum Development", *NACTA Journal*, June 1989.

3. Lavrusik, V., "10 Ways Universities Are Engaging Alumni Using Social Media", July 2009, Retrieved December 2010, from http://mashable.com/2009/07/23/ alumni-social-media/.

4. Shaindin, A., Allen, E., Online Social Networks: A Developing tool for alumni associations, *An Alumni Futures White Paper*, January 2008, Available at: http://www.alumnifutures.com/files/AF_Whitepaper_Online_Networks.pdf.

5. Datamining, Retrieved April 2011, from http://en.wikipedia.org/wiki/Data_ mining.

6. Datamining: What is datamining, Retrieved November 2010, from http:// www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamini ng.htm.

7. Simoudis, E., "Reality check for data mining", *IEEE Expert: Intelligent Systems and Their Applications archive*, Vol. 11(5), 1996.

8. Wu, J., "Business Intelligence: The value in mining data", *DM Review Online*, February 2002.

9. Zhu, X., Davidson, I., "Knowledge Discovery and Data Mining: Challenges and Realities, IGI Global, Hershey, PA, 2007.

10. Squier, L., "What is Data Mining?", DAMA-NCR Meeting, November 2001, Available at: http://www.dama-ncr.org/Programs/ScheduleOfPast.html.

11. Schwartz, J., "13,000 Credit Reports Stolen by Hackers", N.Y. TIMES, May 17, 2002, C5.

12. McDonough, B., “Developers of deep Web mining technology acknowledge concerns about how personal data is used, but say the problem must be handled by those who decide what information goes on the Internet”. *NewFactor Sci:: Tech*, Vol. 5, 2002.

13. CVS Caremark, Retrieved February, 2011 from http://en.wikipidea.org/wiki/ CVS_Caremark.

14. Advantages and Disadvantages of Data Mining, Retrieved December 2010, from http://xiangyun86.wordpress.com/2006/12/05/advantages-disadvantages-of-data-mining.

15. Datamining algorithms (Analysis Services – Datamining), Retrieved March 2011, from http://msdn.microsoft.com/en-us/library/ms175595.aspx.

16. Romero, C., Ventura, S., Espejo, P.G., Hervas, C., “Data Mining Algorithms to Classify Students”, *Proceedings of the First International Conference on Educational Data Mining*, Montreal, Canada, 2008, 8-17.

17. Baker, R.S.J.d. (in press) Data Mining for Education. To appear in McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier.

18. Amershi, S., Conati, C. “Automatic Recognition of Learner Groups in Exploratory Learning Environments”, *Proceedings of ITS 2006, 8th International Conference on Intelligent Tutoring Systems*, 2006, 463-472.

19. Beal, C., Qu, L., and Lee, H. “Classifying learner engagement through integration of multiple data sources”, *Proceedings of the 21st National Conference on Artificial Intelligence*, 2007.

20. Yang, T., "Computational Verb Decision Trees". *International Journal of Computational Cognition* (Yang's Scientific Press), Vol. 4 (4), 2006, 34–46.

21. Decision trees, Retrieved March 2011, from http:// www.d.umn.edu/~rmaclin/ Notes/L03_Decision_Trees.pdf .

22. Weka, Retrieved February 2011, from http://weka.wikispaces.com/Primer.

23. Orange datamining tool, Retrieved February 2011, from http://en.wikipedia.org/wiki/Orange_(software).

24. Rapidminer datamining tool, Retrieved February 2011, from http://en.Wikipedia.org/wiki/RapidMiner.

25. Arceo, A., "The role of student and alumni associations in the democratization process in Spain", *Higher Education in Europe*, Vol. 28(4), 2003, 523-528.

26. Spasić, Ž., Peja, P., "Feedback Information of ALUMNI Association in Quality Assurance System for Higher Education", *FME Transactions*, Vol. 33, 2005, 103-109.

27. Barnard, Z., "Online community portals for enhanced alumni networking", Ph.D dissertation, 2007, Available at http://hdl.handle.net/10210/851.

28. Delavari, N., "Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System", *Proceedings of 6th International Conference on Information Technology Based Higher Education and Training*, F4B/1 - F4B/6, 2005.