ARISTOTLE UNIVERSITY OF THESSALONIKI
SCHOOL OF APPLIED SCIENCES

Christos Melissidis

Discovering Human Relationships in Social Networks
through Grid Computing

COMPUTER SCIENCE DEPARTMENT

MSc THESIS
Discovering Human Relationships in Social Networks through Grid Computing

Supervisors: Athena Vakali, Christina Mpoutsouki
External Supervisors: Frank Wang, Sinning Wu

September 2009

This thesis is submitted in partial fulfilment of the requirements for the degree of Master of Science

© Aristotle University 2009. All rights reserved. No part of this publication may be reproduced without the written permission of the copyright owner.
Acknowledgments

My deepest gratitude goes to my family, my father Ioannis Melissidis and my mother Kalliopi Kolini, for giving birth to me at the first place and supporting me throughout my life. Also to my grandparents, Menelaos Kolinis and Maria Nikolaidou for their love and support.

My sincerest thanks to my professor and supervisor Dr. Athena Vakali from Aristotle University of Thessaloniki, who helped, guided and supported me throughout my studies and efforts until today.

I wish also to thank my supervisor from Cranfield University, professor Frank Wang, for his encouragement and support throughout the duration of the MSc course and the BCS “Grid Computing Now! 2008” Competition.

I am deeply indebted to my colleagues from the Information Technology Center in Aristotle University of Thessaloniki, and especially to my manager Ioannis Salmatzidis who gave me the opportunity to work for 3 and a half years in a cutting-edge IT environment, an experience that opened my mind and allowed me to see life from a totally different point of view.

I would also like to thank all my friends and colleagues in Cranfield and Aristotle University, who made these 2 years an amazing and unique life experience.
Abstract

This thesis focuses on the discovery of human relationships in Social Networking databases by using the Grid Computing technology. Furthermore, it explores the field of targeted advertising in specific groups of people by using the discovered human relationship data. For these purposes, we propose a grid service that searches for people according to criteria that are specified by the user. This service is able to find and present not only individuals that match these criteria, but also people that are related to these individuals in some way. Also, it has the ability to identify people and groups with common interests and activities and therefore, apply targeted advertising techniques through the same system. In order to demonstrate the proposal, a web-based application has been developed. This application is able to simulate the functionality that the proposed grid service would have.

Keywords:
grid computing, human relationships, relationship discovery, social networks, data mining, targeted advertising, grid service.
Contents

Acknowledgements ................................................................. 3
Abstract ................................................................................. 4
Contents ............................................................................... 5
List of Figures .......................................................................... 8

Introduction ............................................................................ 9
Motivation .............................................................................. 10
Thesis aims and objectives ...................................................... 11
  Grid Computing into Business ............................................... 11
  Social Networks and Grid Computing cooperation .................. 11
  Optimizing Targeted Advertising .......................................... 11
  Simplify Databases Distribution ............................................ 12
  Simulation of the idea ............................................................ 12
Thesis outline .......................................................................... 12

Chapter 1 .............................................................................. 13
Thesis Background ................................................................... 13
  1.1 The Grid Computing ....................................................... 13
  1.2 Social Networking Services ............................................ 15
    1.2.1 Types of Social Networks .......................................... 17
    1.2.2 Social Network User Actions ...................................... 18
  1.3 Data Mining .................................................................... 20
    1.3.1 Data Mining and Grid Computing ............................... 21
    1.3.2 Data Mining and Social Networks ............................... 22
  1.4 Targeted Advertising ....................................................... 23
    1.4.1 Content Based and Social Based approaches ............... 23
    1.4.2 Internet-based types of Targeted Advertising .............. 24
    1.4.3 Targeted Advertisement and Human Relationships ....... 25
  1.5 Human Relationships in Software ..................................... 26
  1.6 Grid Services ................................................................. 28
    1.6.1 A Web Service .......................................................... 28
    1.6.2 A Grid Service .......................................................... 29
  1.7 Chapter Conclusion ........................................................ 30
Chapter 2

The Proposal: Social AdGrid

2.1 The Idea
2.2 Information Data Retrieval
2.3 Information Data Storage in the Grid
2.4 The Social AdGrid
2.5 Social AdGrid Proposed Architecture
2.6 Using Social AdGrid
2.7 Chapter Conclusion

Chapter 3

Social AdGrid Simulator v1.5: Demonstration and Architecture

3.1 Simulator’s Architecture
3.2 Homepage
3.3 Register
3.4 Contact Grid
3.5 Service Information
3.6 Log In and Log Out
3.7 My Account
3.8 Grid Request – The Main Component
3.9 Progress
3.10 Search Results
3.11 Advertisement
3.12 Use Case Diagram
3.13 Main Use Scenario
3.14 Chapter Conclusion

Chapter 4

System Development and Code Analysis

4.1 Programming the Simulator
4.2 The Database
4.3 Log In and Log Out
4.4 My Account Data Retrieval
4.5 Grid Request Search Criteria Form
4.6 Search Request Processing Form
4.7 Mining the Database
4.8 Showing Search Results
4.9 Advertising Email Form
Discovering Human Relationships in Social Networks through Grid Computing

Christos Melissidis

4.10 Apply to Join the Grid Service .................................................................73
4.11 Contact Email Retrieval ........................................................................75
4.12 Loading Panel Script ..............................................................................76
4.13 Chapter Conclusion ................................................................................77

Conclusions and Future Work ....................................................................78
Conclusions .....................................................................................................78
Future Work ....................................................................................................80

References ......................................................................................................81

Appendix A .....................................................................................................83
Environment Configuration .............................................................................83
A.1 Apache and MySQL installation ..............................................................83
A.2 PhpMyAdmin ..............................................................................................83
A.3 Notepad++ ..................................................................................................83

Appendix B .....................................................................................................84
Scripts and Widgets .........................................................................................84
B.1 jsProgressBarHandler ..............................................................................84
B.2 Cross Browser Rich Text Editor ..............................................................84
B.3 YUI Loading Panel ....................................................................................84
List of figures

Figure 1.1: The idea of Grid Computing ................................................................. 14
Figure 1.2: A Social Network .................................................................................. 16
Figure 1.3: Data Mining: The core of knowledge discovery process .................... 20
Figure 1.4: Targeted Advertising in Google Search Results ............................... 25
Figure 1.5: The two types of human relationship models in software .................. 27
Figure 1.6: Web Service example .......................................................................... 28
Figure 1.7: Web Service Invocation ...................................................................... 29
Figure 1.8: Grid Services ....................................................................................... 30
Figure 2.1: “Social AdGrid” and “Social Grid In Ltd.” logos .................................. 33
Figure 2.2: Social AdGrid functionality simplified ................................................. 34
Figure 2.3: Social AdGrid proposed architecture .................................................. 35
Figure 3.1: “Social AdGrid Simulator v1.5” Component Architecture .................. 38
Figure 3.2: Social AdGrid Home page ................................................................. 39
Figure 3.3: Log-in form and Apply Now link ......................................................... 40
Figure 3.4: Social AdGrid Registration Form ....................................................... 40
Figure 3.5: Social AdGrid Contact Form for visitors ............................................. 41
Figure 3.6: Social AdGrid Contact Form for registered members ....................... 41
Figure 3.7: Social AdGrid Service Information .................................................... 42
Figure 3.8: Log In Form ......................................................................................... 42
Figure 3.9: Logged in – Log out Form .................................................................. 42
Figure 3.10: Denial of Access for selection “Grid Request” and “My Account” ....... 43
Figure 3.11: My Account ...................................................................................... 44
Figure 3.12: Initializing Grid Service ..................................................................... 46
Figure 3.13: Search Criteria .................................................................................. 46
Figure 3.14: Supported Social Networks .............................................................. 47
Figure 3.15: Human Relationships Discovery ...................................................... 47
Figure 3.16: User provided criteria ....................................................................... 48
Figure 3.17: Processing user’s request ................................................................. 48
Figure 3.18: User accepts and asks for the results ............................................... 49
Figure 3.19: Search Results .................................................................................. 49
Figure 3.20: Advertising through the Social AdGrid grid service ......................... 51
Figure 3.21: Email Text Editor for plain text ....................................................... 52
Figure 3.22: Email Text Editor for HTML language ............................................. 52
Figure 3.23: Use Case Diagram ........................................................................... 54
Figure 3.24: Main user’s actions ......................................................................... 55
Figure 4.1: Database and tables in phpMyAdmin .................................................. 57
Figure 4.2: User “socialadv” in table “users” ....................................................... 58
Figure 4.3: Users in table “social_users” ............................................................. 58
Figure 4.4: Field types and collation .................................................................. 58
Figure 4.5: Data passing through .php files ......................................................... 68
Introduction

Today, the world is experiencing an impressively large wave of increasing information in social networks. As the number of these networks grows, millions of people join and participate in the communities that are behind them. It is hard to imagine the size of the information that is stored in social networks and even harder, to imagine the value of this kind of information.

Millions of users reveal a large part of their personal data when joining a social network. Considering the millions of users and the hundreds of social networking websites, somebody can assume that we are talking about massive amounts of very important and at the same time, fragile information.

Companies and advertising agencies from around the globe are very interested in these data for various reasons. The most important reason is that by using these data, it is very easy to optimally discover the proper target groups to advertise themselves or their products and services. This is a marketing technique called targeted advertising and today it takes several forms on the Internet.

So, it comes to a point that we have to find solutions of how to control all this information and how to effectively distribute it. It is obvious that we need to apply highly demanding data mining, pattern matching and many more techniques and algorithms on huge amounts of data from various and distributed databases.

This is where Grid Computing joins. It provides what it is needed from computational and data resources and is not commonly used in social networking. What this Thesis proposes, is a grid service whose main job is to extract data from several social networks using the techniques mentioned above. The main idea is to provide the system with specific criteria and to request the discovery of human relationships and people who match these criteria. By using this service, the client can find the target group that is interested in and apply targeted advertising through this grid service. We named this grid service “Social AdGrid” and we simulated it in a very detailed application, in order to present the proposed idea.

It is obvious that several challenges arise from this new proposal, such as privacy and data collection. However, we strongly believe that the combination of the cited technologies could be a step ahead in this area.
Motivation

It is known to most computer scientists or engineers that until today, Grid Technology has not been used or applied in the advertising market. On the contrary, this technology is mainly used for scientific and research purposes. Initially, the proposal was thought as a way for grid computing to bring profits and to enter vigorously the business area. This could be accomplished, if something not only effective, but useful and attracting could be proposed.

It is also known that social networks are experiencing a significant growth and this will continue at least in the near future. Also, today million of dollars are spent in order to achieve the most effective type of advertising. Advertising agencies have realized that internet is a very good advertising tool and that they should invest on that. There are a lot of internet advertising techniques, but the question arises for which one is the most effective.

Every person is an individual with specific needs and tastes. Also, it has been proved that in real life, people with same interests and needs tend to socialise and create relationships. This is also true in social networks. In the era of social networking, people tend to share this personal information by either declaring them directly in their social networks profile, or by joining specific online social groups.

Human relationships can tell us a lot about how people think, act, live and buy. This makes human relationships information very valuable to advertising agencies, in order to create marketing target groups and to apply the so called targeted advertising.

But this kind of information is huge and it needs more than today’s common computing infrastructure in order to be discovered. As mentioned, techniques such as data mining need to be used in order to deal with the information discovery in this amount of data. The opportunity of using grid computing in order to discover that knowledge, gave us the initial idea of building a service that could meet these needs, by putting grid technology into the advertising market at the same time. It is obvious that we are talking about a large amount of distributed data and the need of proposing a unique virtual database was also necessary. What we want to achieve in this Thesis, is to propose a new idea and with a simulation, to give the user a better understanding of how this would work.
Thesis Aims and Objectives

In this chapter we will mention the main aims and objectives of this Thesis, as well as the reasons that led us to come up with the proposed idea. The goal of this project is not only to give an explanation of our idea, but also to provide a hands-on experience of how the proposed model would work in real life.

Grid Computing into Advertising Business

The main aim of this thesis is, as mentioned, to introduce Grid Computing into the advertising business area. So far, Grid technology has been used mainly for scientific and research purposes by various organizations such as the European Organization for Nuclear Research (CERN), NASA etc. The proposal is a “pay-by-results” advertising Grid Service that charges the user, depending on the consumption made, such as the resources used and the time employed by the system.

Social Networks and Grid Computing cooperation

It is not common for social networks and grid computing to cooperate and collaborate. There are some social projects that involve grid technology, but nothing has been applied or developed. This is one of the main reasons why to combine these two computer science fields. Other reasons, such as the huge amounts of data included in social network databases, as well as the computational power need for the application of demanding techniques and algorithms, play an extremely important role.

Optimizing Targeted Advertising

Internet advertising is a field of marketing that is now being explored and researched. Advertising agencies are more and more interested in this field. With the growth of social networks, they are trying to find out new ways of advertising using the data derived from these networking services. Targeted advertising seems an advertising technique that can be used very effectively in these networks, by filtering and finding specific individuals and advertise to them directly or with indirect ways. One of our most important objectives is to combine the two computer science technologies mentioned before, in order to optimize this type of advertising.
Simplify Databases Distribution

There are at least double as many social databases, as the number of social networks today. That makes the number of social databases to tens of hundreds. It is obvious that is very difficult to retrieve data from such a large number of geographically distributed databases. The main purpose is to create a unique virtual database that virtually connects all the other databases. This would make data mining and other techniques easier for the system that will manage this virtual database. It immediately came up that we are talking about Grid Computing.

Simulation of the Idea

In order to present the proposal, it was necessary to develop something that would demonstrate this idea. A Grid Service of that kind would be impossible to be implemented by one person in half a year, but a simulator of this Grid Service would perfectly serve our purposes.

Thesis Outline

The structure of this thesis project is as follows:

- Chapter 1 gives an overview of the technologies that are involved in this thesis, such as Social Networks, Grid Computing and Targeted Advertising.
- Chapter 2 introduces our main thesis idea. It describes what exactly the proposed service is. Simple diagrams and images are given in order to softly introduce the reader to the proposal.
- Chapter 3 demonstrates the simulator that has been developed in order to present the proposed grid service. This manual-like chapter provides screenshots and images for demonstration purposes. The reader can have a complete idea of how the proposed grid service works.
- Chapter 4 explains the most important parts of the code that has been programmed and used in order to create the grid service simulator. The reader is able to see the actual code and its descriptions.
- Chapter Conclusions and Future Work presents the summary of the work done in this thesis and gives suggestions for possible future improvements.
Chapter 1

Thesis Background

This chapter introduces the technologies and science fields that the reader needs to know in order to understand, and to be able to follow, this thesis document. The project is based on three main computer science technologies: Grid Computing, Data Mining and Social Networks. It is also based on marketing techniques and methods, such as Targeted Advertising. In the following paragraphs, terms and definitions for the mentioned science fields and methods are given.

1.1 The Grid Computing

Grid Computing is a very promising technology and it has been said by several gurus of computer science that it is the “the next big thing”. What it does is using several computers or computing resources at the same time, in order to solve a single problem. The problems that are being solved are usually scientific or technical problems that require great computing power or access to very large amounts of data.

Grid Computing refers to a distributed technology. This means that the resources that belong to a Grid Network are geographically allocated in different places. This is the main reason that grid technology is distinguished from conventional distributed computing systems.

The term for Grid Computing is that the “The Grid” is a number of computers or data resources connected together and working in parallel, in order to solve smaller problems that are part of a larger problem which has been split by the system. The software that is being used by Grid Computing is able to farm out the pieces of a program or problem to several computers, sometimes in numbers of thousands.
Grid Computing is a very valuable and important technology for several reasons.

- First of all, it has the ability to use more effectively the given computer resources.
- Secondly, it can be used as a way to solve problems that cannot be approached without the use of very large amounts of computational power.
- Finally, this technology suggests that the resources of many computers can be harnessed cooperatively and managed, in order to achieve a common objective.

Grid Computing has been applied in many fields of the industry and science. Many grid projects have been funded by the European Commission through the framework programme, such as BEin-GRID (Business Experiments in Grid) and EGEE (Enabling Grids for E-science).

One of the most important centres for the development of Grid is CERN (European Organization for Nuclear Research). CERN uses Grid Computing for several purposes such as processing data of extra terrestrial intelligence and in the development of LHC (Large Hadron Collider). An example that is widely known is SETI @ Home (Search for Extraterrestrial Intelligence). In this project, many thousands of users created a distributed network and shared for free the resources of their personal home computers in order to analyze data of outer space signal searching.
As mentioned, Grid has also been applied in several science fields. In Biology, scientists have used grid computing for the simulation of thousands of molecular drug candidates. Earth scientists and geologists are using it to track the ozone levels using satellite inspection. High energy physicists are using grid infrastructures in order to understand better our universe. Engineers are using grid technology to study alternative fuels as fusion energy and artists to create very complex animations and graphics for films.

Globus Toolkit is an example of an open source software that is being used in order to create and build Grid Systems and Grid Applications. Through Globus Toolkit, people are given the ability to share computational power, databases and other tools without institutional and geographical boundaries. This software contains services and libraries for specific purposes, such as resource discovery, monitoring and management. It also includes other several components, such as for security and portability. [14]

1.2 Social Networking Services

Social Networking Services are transforming the way that people use Internet and interact with each other. Today, these services are developing rapidly as technology is advancing and they are on the rise globally. While the number of social network users is increasing faster and faster, more and more new services are being launched. It is obvious that the term “Social Networks” is swiftly expanding. But what is a Social Networking Service and what are the forms that these services take?

The term “Social Networking Services” is a wider term for “Social Networks” and refers to a very wide range of tools and practices that are being rapidly developed today. Social Networking Services can be generally defined as “internet based social structures that are designed to offer communication, collaboration and content sharing between their users”. [11]
More specifically according to Danah M. Boyd and Nicole B. Ellison:

“We define social network sites as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site.” [2]

Social Networking services allow their registered users to build their own social profiles and furthermore to manage them. Also they provide the users with services like events, groups and member relationships. Additionally, they offer to both users and groups privacy settings and restrictions, as well as profile access management, usually with specific access levels. Frequently, these settings vary from one social network to another, but there are some general settings that can be mentioned:

- **Profile Information Visibility Restriction:** Service registered and certified members, people on user’s contact list, specific user groups.
• **Information Privacy:** Information viewable only to public users and specific users selected by the account holder.

• **Public Profile Information:** Information viewable to not registered members and website visitors.

With all these privacy settings combination, the user is able to manage his/her online profile appearance to friends, groups or public.

Social Networking Websites also vary in their functionality as well as in the variety of tools which they offer their users. Though, Boyd and Ellison define social networking websites to have three main common elements: a web page member’s profile, the ability of the user to add other social members in his/her contact list and the interaction between members of contact lists. [2]

1.2.1 **Types of Social Networks**

There are a lot of theories about the types of Social Networks. However, in this thesis we consider two main formats or types of social networking websites. The ones focused on the user’s profile (ego-centric) and the ones focused on the content (object-centric). In simple words, ego-centric social networks place the individual as the core of the social networking experience. Such networks are Facebook, LinkedIn and Friendster. On the other hand, object-centric social networks place at the centre of the social network a non-ego element, such as photographs, videos and hyperlinks. Such networks are Flickr, Dopplr and Digg. [9]

Despite the mentioned main categories, more types of social networks can be further distinguished. The most important are described shortly:

• **Profile-Based Social Networks:** Usually, they are focused on member’s profile pages that consist of personal information about an individual member. Such information can be personal details, hobbies, interests, and member’s pictures. In this kind of social networks, member can interact with each other by exchanging information, embedding links, wall messaging and more. Facebook and my Space are very good examples of this category.
• **Content-Based Social Networks:** In this type of Social Networks, the user’s profile plays a secondary role. More specific, in these object-centric social networks the most important thing is the content. Members can see the content that other users have uploaded. Music, photos, books and videos are the main content types of these networks. Some very good examples are YouTube, Shelfari, Last.fm and Flickr.

• **Multi-User Virtual Environments Social Networks:** In these social networks, the user’s profile is a virtual representation of himself/herself. They interact with other virtual members of the network and they can build, control and move their character or even fight virtually between each other. Two very famous examples of these networks are Second Life and the very famous World of Warcraft.

• **Micro-blogging Social Networks:** In these social networks, the user is allowed to publish short messages. These messages can be status updates and emotional or mood changes. Other users can see these messages and comment on these updates. Such networks are Twitter and Jaiku.

### 1.2.2 Social Network User Actions

But what can a user do in a social network? Because of the many kinds of social networks, we will try to focus on the most important actions that a user can perform in a social network. Such actions include:

• **Gathering contacts and developing friendships.**  
  People in a social network can communicate and socialize with their contacts or groups of people.

• **Personal representation.**  
  People are joining social networks in order to present themselves not only to their contacts, but to the general public as well.

• **Finding and viewing information or content.**  
  This is one of the most important actions of the social network users. People tend to like keeping themselves updated with other’s people status changes and updates. Members can also view other user’s uploaded content, such as pictures, videos and events.
• **Building and managing profiles.**
  In every social network, a profile is a different set of tools. Profiles allow members to show and manage a general “image” of themselves to other social network’s users. In a user profile, several things can be viewed. Such things are wall posts, favourite videos or music, avatars and personal information.

• **Uploading, managing and sharing content.**
  In the majority of social networks, users have the ability of uploading their own content. This content can be, as mentioned, pictures, videos or music. It can be visible by their friends, groups or the general public. Content can also be third party, such as articles, videos or pictures that are hosted in several websites, or even other social networks.

• **Message exchanging.**
  Users are given the ability of messaging, private or public. This messaging can take mainly two forms. The first is instant messaging (known as chatting) through a website component and the second one is email based.

• **Communicating with other people.**
  This is the most important action that is offered to the users of social networking services. This is actually why they are called “social”. People can make discussions, comment, share and communicate.

Several types of Social Networks offer their users a variety of actions. They allow them to socialize and communicate with other people in many ways. They also allow the sharing of various content. Millions of people are experiencing the use of social networking websites with thousands joining every day. Furthermore, the uploaded content is increasing dramatically. And the question arises, where is all this information stored and what techniques can be used in order extract it? More importantly, how can this information be used?
1.3 **Data Mining**

Nowadays, we are experiencing an unexpected and more importantly an unprecedented information growth, in a variety of science and knowledge sectors. Many researchers are talking about an “Information explosion” phenomenon, by describing with this term the increasing amount of data that are published and the way that this is affecting our today’s society.

Companies are producing and storing huge volumes of data. Also, scientific organizations and science centres store and access very large databases with complex and valuable data, such as biological or astronomical and economical data. The more complex are these data, the more difficult and time consuming is their storage and further analysis.

![Data Mining: The core of knowledge discovery process](image)

According to Fayyad, Piatetsky-Shapiro and Smyth, “Data Mining is a non-trivial process for identifying valid, novel, potentially useful and understandable patterns in large volumes of data”. The information that is being extracted can be used in business applications, such as customer purchase behaviour. [20]

Data Mining is a very complex process and is today one of the most important fields of computer science, due to the importance and the amount of information today.
Briefly, the three main dimensions of data mining complexity are the ones that follow. [1]

- **Data Mining Tasks:** The data mining process includes many non-trivial tasks. Such tasks are data pre-processing, model validating and presenting results.

- **Data Volume:** Today, the applications that apply data mining are required to analyze rapidly growing amounts of data, containing tens of millions of entries and consisting Petabytes of data.

- **Data Complexity:** Complex data structures are being created by capturing complex analyzed phenomena. Also, data are more and more geographically distributed and cannot be gathered centrally for several reasons, such as privacy, legal and security reasons. [7]

### 1.3.1 Data Mining and Grid Computing

As mentioned, databases today tend to be more and more distributed. As a result, a change in the term data mining is taking place. Data mining is becoming “distributed data mining” and as we move slowly to the grid era, data mining is becoming grid-like or even grid-based. Today, computer scientists are talking about data mining grids, which seek to maximize the effectiveness and the efficiency of the entire data mining process, with a trade-off between distributed processing and data centralization. [1]

The new version of data mining, called “grid-based data mining”, would allow the following. [3]

- To distribute demanding data analysis processes to a big number of distributed resources.
- To develop new techniques and algorithms for local data processing with result of avoiding transmissions, thus privacy issues.
- The researching and finding solutions for data mining problems.

Until today, the cooperation and combination of data mining and grid computing was successful and there plenty of examples to prove that. These two technologies have been married in a variety of sectors, such as molecular biology,
weather modelling, environmental problems and nuclear experiments. But could this combination be applied in a very new and challenging area such as the Social Networking?

1.3.2 Data Mining and Social Networks

It is easy to imagine for most of us, the amount of information and the value of this information that is stored in a social network. Millions of people trust their personal information and upload their content in these social networks. Petabytes of complex data are stored in this kind of social databases. Of course, it can immediately be detected the need of security for this information and the importance of the privacy protection. The information that is hidden in these databases is extremely valuable and it contains very sensitive data. If this kind of data is used in the wrong way, people could face serious problems and could witness a variety of frauds. As expected, many companies are interested in this data, for various reasons, such as advertising and product promotion.

Today, little work has been done regarding external data mining from social networking databases. Mining in social data is very restricted and can be done through authorized services and applications. On the other hand, not all the social data are extracted through applications. A large amount of data is available for free collection on the internet. With techniques such as text mining, valuable information can also be retrieved.


1.4 Targeted Advertising

Businesses have recognized long ago the importance of targeting a small portion of customers to advertise and the effectiveness of it, for two main reasons. [19]

- The first reason is that the volume of products or services information that is available to the customers is increasing. This means that it is needed to help customers find the desirable information for the products or services they want.
- The second reason is that the understanding of the needs of current or future customers is a very important component of customer relationship management.

“Targeted advertising is a type of advertising whereby advertisements are placed so as to reach consumers based on various traits such as demographics, purchase history, or observed behaviour.” [18]

The traditional targeted advertising is the manual analysis of a historical database that contains previous customer’s transactions and features. It can be accomplished with statistical tools. Then, the advertiser is able to extract a customer list and there is a great possibility for these customers to respond to the advertisement of the specific product.

Today, the arrival of many new technologies has brought what is called “automatic customer identifying tools”. These tools are able to spot customers according to their preferences that are obtained by transaction records, cookies, internet blogs and more.

1.4.1 Content Based and Social Based approaches

The traditional targeted advertising technique was the so-called “content-based approach”. In this technique the products are marked with a set of specific characteristics and then customer’s preferences are represented by a same set of characteristics. These approaches are able to detect customer target groups whose interests are very similar to product’s specifications. Of course, these approaches are not appropriate for kinds of products that are not electronically available, because they are based on text extraction or text mining. Additionally, because the customer characteristics are retrieved from his/her history purchase, it is not effective to advertise products to new customers. [21], [22]
A later type of recommendation technique is the “social-based approach”. This approach searches for relevance between customers. This is accomplished by observing their ratings to specific products. Customers with relevance to the target customer are found and products are being advertised to them. This approach has been proved useful in many applications. Despite that, it appears to have some limitations, such as that new unrated products cannot be advertised and also, customers that have not rated any products cannot be advertisement targets. [21], [22]

1.4.2 Internet-based types of target advertising

Talking about targeted advertising on the Internet, there are two main types, as found in most of the bibliography:

- **Behavioural Targeting:** “A technique that is used by online publishers and online advertisers to increase the effectiveness of their campaigns.” [16] This technique is using data that are collected from the customer’s web browser’s behaviour. This kind of data can be visited web pages or executed searches in various search engines. In this way, specific advertisements can be selected and shown to specific customers. Some of the advertising targeting systems that use behavioural advertising are: AdLINK 360, Boomerang, ValueClick and Predicta BT.

- **Contextual Advertising:** “A type of targeted advertising for advertisements appearing on websites or other media. The advertisements are selected and served by automatic systems and are based on the content that is displayed to the user.” [17]

A system that uses the second type of advertising is able to scan the text of the specific website. By doing that, it is able to identify keywords or tags and the content’s category or sector. After that, it places the proper advertisements in the webpage that the user is viewing, for example as pop-up ads. Search engines are using this type of targeted advertisement in order to display advertisements on their result pages. Figure 1.4 shows that very clearly.
Targeted Advertisement and Human Relationships

Human relationships play a very important role in marketing and especially in targeted advertising. In real life, a customer that is about to buy a product is very often influenced by his or her family, relatives, friends, colleagues etc. Furthermore, people with same interests tend to socialize. This means that if a customer is found as a target, there is a great possibility that he/she shares interests with his/her friends or people that he/she knows.

In order for the reader to understand, a specific example is given. Emails are sent usually between people that are related in some way (friendship, family, colleagues etc). The Hotmail free email service had grown from zero user accounts to 12 million accounts in just 18 months. This happened because the company included a promotional message for the service in every email sent using it. But the main reason of this growth was that people were influenced by the behaviour of their related contacts. [23]
Today, it is understood that customers with strong ties tend to show similar purchasing behaviour, because of the mutual and bidirectional influences. It’s a very promising idea to take actions like:

- Identifying groups and subgroups.
- Finding related people and hence with common interests and activities.
- Finding target groups and portions of customers for advertising, through data mining in social networking services.

As mentioned, we realize that social networks have very important and valuable information that can be used in marketing and more specifically in advertising. In fact, the effectiveness of the targeted advertising can exceed every method’s efficiency until today.

This is the main idea of this thesis and it was inspired from this exact fact. The reason was that in Social Networks, not only a possible customer target shares interests with friends, but also with groups of people. For example, if someone likes ice hockey, he/she will probably join a group in Facebook named “Ice Hockey Fans”. Hence, there is a great possibility that people from this group will be his/her friends. Immediately, it becomes obvious that a small target group has been discovered. The people of this target group will most probably be positive in an advertisement about low-cost hockey tickets. As mentioned, grid computing is the mediate technology that is used in order to extract these exact data from a number of databases.

1.5 Human Relationships in Software

In this thesis, we accept the relationship pattern that is described in this paragraph. In general, there can be distinguished two main ways of modelling human relationships in software and furthermore in social networks. These models have been identified after the creation of social networking services. [10]

- The first type of model is the “Asymmetric model”. In this model a social user ‘A’ is able to follow another user ‘B’, without the user ‘B’ following ‘A’ back. It is mainly a one-way relationship and it may be or may not be mutual. The most famous social network that uses the “asymmetric model” is Twitter.
• The second model type is the “Symmetric model”. Between the user ‘A’ and ‘B’ there is a two-way relationship. This means that if user ‘A’ adds ‘B’ as a friend, ‘B’ has to do the same. As a result, users of these social network types have always 1-1 relationships with their friends. The most famous social network that follows the “symmetric model” is Facebook and LinkedIn.

![Image: The two types of human relationship models in software](image)

Figure 1.5: The two types of human relationship models in software

The “asymmetric” model allows four types of relationships in comparison with the “symmetric” one, which offers only two (friends or not friends). The asymmetric types are:

1. People follow user ‘A’, but ‘A’ does not follow them back.
2. People do not follow ‘A’, but ‘A’ follows them.
3. Both people and ‘A’ follow each other (relationship known as friends).
4. Neither of them follows each other.

Both models have advantages and disadvantages, but until today there is no clear evidence of which one will prevail in the near future. [10]
1.6 Grid Services

In this paragraph, it is given a brief explanation of the term Grid Service. But since Web Services are the basis for the Grid Services, it is fundamental to understand the web services architecture. So, what exactly is a web service?

1.6.1 A Web Service

Very simply, web services are a distributed technology, such as CORBA and RMI, which allow the creation of client-server applications. In order to understand better, a simple example is given below.

A chain of book shops is all around U.K. but the main catalogue of the products is available only in the database that is placed in the company’s central offices in London. The others shops though, should be able to access this catalogue of products. This can be done by a web service, as seen in figure 1.6.

![Web Service example](image)

**Figure 1.6: Web Service example**

Web Services have some advantages, over other similar technologies:

- They provide interoperability between various applications that run on different platforms.
- They use open standards and protocols. These data formats and protocols are mainly text-based.
- They allow services and software from several locations and various companies to easily collaborate, in order to provide a unique service.
- They are able to function through the most common firewall security measures, by utilizing HTTP. Changes to the filtering rules of the firewalls are not usually required.
In the figure 1.7 there is a complete web service invocation. The steps of the procedure are numbered from one to five.

![Figure 1.7: Web Service Invocation](image)

### 1.6.2 A Grid Service

As mentioned, web services are the proper technology for creating web-based applications, with loosely coupled clients and services. This makes this technology the optimal choice to build the grid-based applications. Though, in web services there are certain limitations and plain web services cannot be used in order to build a grid application. So generally: “Grid Services are web services with improved characteristics and services”. [13] As seen in figure 1.8 a grid service is defined by the OGSA (Open Grid Services Architecture) and is specified by the OGSI (Open Grid Services Infrastructure). Also, as mentioned, a grid service is an expansion of a web service.

Briefly, the purpose of the OGSA, which has been developed by the Global Grid Forum, is to define a common and standard architecture for grid based applications. Its goal is to create and specify standards for all the services that somebody can find in a specific grid application, such as security services, resource management services and more. Because OGSA cannot describe all these services
with many details, OGSI was created again by the Global Grid Forum in order to provide detailed specifications of how a grid service works.

The Globus Toolkit, as mentioned in paragraph 1.1, is a software kit developed by the “The Globus Alliance”. Its tools can be used in order to create grid-based applications. It also includes an implementation of the OGSI, as seen in the last figure. It also includes programs, utilities and services that some of them are built on top of the OGSI and are named Web Services (WS) components. Those who are not built on top of the OGSI are named pre-WS components. [14]

1.7 Chapter Conclusion

In this chapter, technologies like Grid Computing, Social Networks, Targeted Advertising and Grid Service have been analyzed. It is believed that the reader is equipped with all the necessary literature knowledge of the sectors, methods and techniques that have been used and that have inspired us for the proposal. In the next chapter, it is presented the main idea of this thesis and the proposed grid service.
Chapter 2

The Proposal: Social AdGrid

In this chapter, we present our main thesis idea. We will stay focused on the theoretical understanding of the proposal and in the next chapters we will move on to the demonstration and the implementation. It is explained how the information can be retrieved and stored in the grid and simple graphs and diagrams are provided in order to demonstrate the main concept.

2.1 The Idea

As mentioned in previous chapters, the internet today, and more specifically social networks, can be an extremely fruitful place for advertising. Techniques such as targeted advertising are being applied, but still there are a lot of things to be discovered and these methods are still in their infancy. This was one of the main motivations for coming up with this idea. The proposal is using and focusing on four main science fields:

- Grid Computing
- Social Networks
- Targeted Advertising
- Human Relationships

Briefly, the main concept is to apply very demanding data mining, pattern matching and other techniques in the databases of social networking services, by using the grid technology. With the help of this technology we are able to find, reveal and extract human relationships and create customer target groups, so that we can later apply targeted advertising techniques to these target groups in order to advertise products and services. From now on, these people that we advertise to are called “customers”, for comfort in referencing. It is important to know that the proposal has been designed as a grid service. By reading this paragraph, it seems a little confusing, but it is explained in detail later on. Before getting into the proposal, it is considered important to explain how this data is gathered by the Grid.
2.2 Information Data Retrieval

Most of us already know that today millions of users place their data in social networking services like Facebook, LinkedIn, MySpace, Twitter and many more. Applications have been developed for these networks, which have access to the user’s personal data. In order for the reader to understand better, an example is given. In facebook, there is an application called “Birthday Reminder”. When the user installs this application, he/she accepts the terms of use and immediately the application gains access to a part of his/hers profile data.

In this proposal, it is assumed that in this way, we can access and retrieve user data from social networks. Of course, there are plenty of ways for retrieving data from social networks, but it need to be kept simple and it is assumed that the system uses this type of information retrieval. Hence, in order to gather data, it is needed a number of social applications like the one in the example.

As soon as these applications are running, the data collection has started and information is already being pumped. In the application’s terms of use, the user of the social network is informed that his/her information is used by our system (grid), in order to release him from spamming-like advertisements and to perform an optimal advertising that would fit his/her needs.

After that, the social user is using the application by his/her own will, and he/she is providing the system with very valuable information. As it can be ascertained, if this application would work optimally, there could be tens of thousands of users that would join the system. This information is stored in several databases, but we will get into this more deeply in the next paragraph.
2.3 Information Data storage in the Grid

The Grid is gathering information at all times. This information needs to be stored in several social databases and usually, these databases are geographically distributed. For example, user information that is gathered from United States will be stored in a database placed in the same country.

All these distributed databases are forming a virtual database in our grid. The grid is responsible for managing these databases and connecting our grid service with this virtual database. Now that we have explained how the user information is collected and stored into the grid, we are able to present the proposal.

2.4 The Social AdGrid

As mentioned in previous paragraphs, this system is designed as a grid service. In chapter 1 it was explained that a grid service is a web service with improved characteristics. We named the proposed grid service “Social AdGrid” and we assumed that this grid service is offered by a company named “Social Grid In Ltd.”. Also, we named the Grid that runs behind this service “Social Grid”. From now on, they will be referenced with these names.

The users that can use this service are advertising agencies and companies that want to advertise their services or their products. These agencies must be registered in order to use this service. Once they join, they are able to search for people and human relationships with specific criteria. These criteria can be personal information, such as

Figure 2.1: “Social AdGrid” and “Social Grid In Ltd.” Logos
age, sex, country etc., but also interests, hobbies, education level and more. People that are related to the target customer can also be shown in the search results if the user wishes so. After the user receives the results, he/she can advertise to the users through the Social AdGrid grid service directly.

A simple diagram is shown in figure 2.2. This diagram explains the grid service’s functionality in simple terms.

![Diagram](image.png)

**Figure 2.2: Social AdGrid functionality simplified**

In the above diagram we can distinguish the users of the grid service who are the advertising agencies. Also, we can see how the grid service interacts with the databases of the various social networks. The most important thing that is derived from the above diagram is that the advertising agencies can advertise through the grid service.

In this proposal, that there are some privacy issues. When we first came up with the idea, everything seemed flawless. But when we had to face the point where the advertiser would take the customer’s data and advertise, something was wrong. According to the private data protection terms, the customer’s privacy was violated. Furthermore, it is not possible to control the agency’s future behaviour. The agency could use the emails from derived from the grid service for spamming or even worse.
Selling email address lists to other agencies is a very profitable business today. So, we realized that we had to protect these private data. In this way, not only the user’s privacy is protected, but also the user’s trust can be earned and a reliable service can be built.

The solution for this problem was that the advertising agency only had the option to advertise its products or services through the same grid service. The agency is given the tools to advertise to a list or to individual customers, after the return of the results and while is logged-in into the system. In the next chapter the abilities of the system are examined in depth.

2.5 Social AdGrid Proposed Architecture

In this paragraph, it is explained the architecture of the grid service and how it works behind the user interface, in simple terms. In figure 2.3 there is a diagram for the proposed architecture of the system. It is presented it as simple as possible, because as mentioned, the main purpose is to brief the user in the idea.
As seen in the figure, several techniques are being applied in the databases of social networking services, such as data mining, pattern matching and machine learning. The Social Grid is responsible for saving (or indexing) all the data derived from these techniques and tools, into a virtual database. As mentioned before, the grid service is using this virtual database in order to connect all the distributed databases of the grid and make data mining an easier procedure. The virtual database contains various data like people profiles, human relationships and other information.

The Social AdGrid grid service has many functions. By looking at the figure 2.3 there can be seen some of them. The user of the service (advertising agency) can do the following:

- Log-in and Log-out of the grid service website
- View and Manage his/her account Information
- Upgrade account to Premium
- Charge the company’s bank account
- Apply for registration in the grid service
- Make a Grid Request (main function)
- Advertise through the grid service tools
- Contact a list of customers or individuals
- Contact support
- Unregister from the grid service

2.6 Using Social AdGrid

In this paragraph it is given an explanation of who is using the proposed Grid Service and for what reasons. The best way to explain is with an example. Let’s assume that the sports equipment company Adidas wishes to advertise one of its new products. The new product is a very flexible and resistant pair of shoes for trekking and soft climbing. This company hires a third advertising agency that will undertake the advertising mission. Among the other marketing types of advertising the agency chooses to advertise this product through the internet, because of its effectiveness in the young target groups.
A very good way of advertising on the web is through Social Networking websites. Advertisements are placed in user profiles depending on the personal data and information. So the agency decides to use this promising targeted advertising technique but with the highest success ratio possible. This can be accomplished if the advertisement appears only to users with specific preferences.

Optimally, it would be best if the advertisement is sent to people with hobbies like trekking and climbing. Then, there would be a high possibility that these people would be very interested in the new offer. Moreover, as mentioned in the literature review (see paragraph 1.4), people with common interests tend to socialize and create groups. This is very common in social networks. Hence, it would be even better if human relationships and user groups could be discovered.

This is where Social AdGrid comes in. Agencies like that could use this grid service in order to optimally discover people with specific criteria such as hobbies and interests. Moreover, human relationships and social groups can also be discovered by the same service. What the agency does, is logging in the service and inserts the specific criteria into the provided search engine. In our case, let’s assume that the company inserts hobbies like trekking and climbing, in Europe and in the age range of 23-29. The selected social network to be searched may vary. The request has been placed and the company is waiting for the results. After the request in the grid, the agency is automatically charged and is given the ability to advertise through the service to the people that have been found.

2.7 Chapter Conclusion

This chapter simply described the main proposal. In order to demonstrate the functionality and the use of this proposal, it was very important to develop a simulation of the grid service. The next chapter is dedicated for the presentation of the simulator that has been developed for this grid service demonstration.
Chapter 3

Social AdGrid Simulator v1.5: Demonstration and Architecture

In this chapter, we present the simulator that we developed for our thesis proposal. The grid service simulator functions are presented with screenshots and images, in order for the reader to understand better the whole concept of the idea. The simulator is also available online for testing. The important parts of the code will be explained in the next chapter. It is important to know that this chapter follows a general “service use” case study.

3.1 Simulator’s Architecture

The simulator has been built as a website. The parts of the simulator are shown in detail in this chapter and in the next one. Before presenting the components one by one, it is needed to present the simulator’s architecture. In the figure 3.1 there can be seen all the components of the simulator and how they interact with each other.

The main one is the Homepage component. This is the page that the user (visitor or register member) views when visits the Social AdGrid grid service. He/she can view this page whether is logged in or not. As seen from the figure, the user can log in (Log In – Log Out component), contact the grid (company.php) or see the Service Information (profile.php). Of course, it is given the user the ability to register to the service with the Register component (apply.php).

If the user is already a member, he/she can log in (logged_in.php) and after that, more choices are becoming available. The user is able to see his/her account with the Account component (account.php) or make a customer search with the Grid Request component (grid_request.php). These choices had restricted access and were not accessible before logging in. Of course there is always the option to log out (logged_in.php). The Account component is communicating with the database (grid_ws_users.sql – table: “users”) to print the user information. Also, it is able to update information in the database, as it is explained later.
In the Grid Request component, the user is specifying the search criteria. After submitting his request, the simulator is loading the Progress component (progress.php). In this one, the searching progress process is being simulated and the user is informed for the charge of the search.

After the simulation of the searching progress, the results appear to the user with the Search Results component (search_grid.php). This component is requesting data from the database and printing them to the screen. The user receives a table with the results and by clicking on specific links he/she is forwarded to the Advertisement component (send_email.php). This is the last one and the user is able to use it more than one times in one search. In the following figure, we are able to see the components that have just been mentioned, as well as the relations between them.

![Component Architecture Diagram](image)

Figure 3.1: “Social AdGrid Simulator v1.5” Component Architecture
3.2 Homepage

In the figure 3.1 appears the homepage of the grid service. The label shows that this has been designed as a simulator, and the version of it can be seen in the upper right corner. The purpose of the homepage is to give the visitor a brief idea of what this grid service is about and of course to attract new users. The visitor can see the main menu of the website, the Log in form and the contact information. Furthermore, some information about the service is available.

Figure 3.2: Social AdGrid Home Page
Except from the main menu, the visitor has the ability to use another menu, in the top of the website. From the main menu, a visitor can only use the selections: “Home”, “Service Information” and “Contact Grid”. He/she also has the ability to join the grid service by clicking the “Apply Now!” link in the login form. If the visitor is a member, he/she can log in by providing his/her username and password, in the login form. In figure 3.3 there is a screenshot with this form.

3.3 Register

If the user chooses to apply for joining the service, a form will appear, like the one seen in figure 3.4. In this form the user must provide the system with some details of the company and the holder of the account. In this system, every account is not associated with a person but with a company. So the details that have to be provided are mainly the company’s details. As mentioned in previous paragraphs, the users of this service are advertising agencies or companies that want to advertise their products or services. Fields like address, email and reasons for joining are also provided.

Figure 3.3: Log-in form and Apply Now link

Figure 3.4: Social AdGrid Registration Form
3.4 Contact Grid

The visitor of the grid service, or even the registered member of the service, is given the ability to contact the support team. With the form that is shown in the figure 3.5, the user can submit his query and his email address, if he/she is a visitor. As seen in figure 3.6, if the user is a registered member and is logged in, the email is automatically filled.

![Figure 3.5: Social AdGrid Contact Form for visitors](image1)

![Figure 3.6: Social AdGrid Contact Form for registered members](image2)
3.5 Service Information

In the “Service Information” subpage, the user (visitor or registered member) is able to see some information about the grid service and the possible users of that service. In figure 3.7 we can see how this page looks like.

3.6 Log In and Log Out

If a user is a registered member of the grid service, he/she can log in by using his/her username and password that have been provided by the service. In figure 3.8 it is shown the log in form that the user can use and in 3.9 the form after logging in.
If an unregistered user tries to click on selections that are only accessible by registered members, the service blocks it and the proper message is shown. In figure 3.10 it is shown the screen that appears if an unregistered user tries to click on “Grid Request” or “My Account” link.

![Figure 3.10: Denial of Access for selection “Grid Request” and “My Account”](image)

3.7 **My Account**

After the user logs in the service, the website changes and it offers him the access to previous restricted areas such as “My Account”, seen in figure 3.11. In this area, the user, which as mentioned is an advertising agency, has the ability to view the company’s account. Also, the company is given the ability to upgrade its account to Premium Account and make a request to delete its account.

Let’s assume that there is a registered user that refers to an advertising agency. This company has been successfully registered before and has used the grid service several times. The username of the account is socialadv and the company’s
name is Social Ad Ltd. A variety of information is available to the user of the service, such as the account holder, email, bank account, website and more.

![Figure 3.11: My Account](image-url)
A very important part of the grid service, is the “Total Requests” and the “Total Charge” that appear in the profile page. The use of this grid service is not free and every time the user makes a grid request, the company’s bank account is charged with a specific amount of money. This amount of money depends on the search criteria and the time consumed for that search by the system, but this is explained in detail later. The user is able to see the total charge, which is the total amount of money, that the company has been accepted to get charged until today. The total requests that the user has performed until today, also appear on his/her account. Furthermore, the user has the ability to provide a logo. As seen in figure 3.11, the company Social Ad Ltd. has given its logo and that appears in the company’s profile and to the other users of the service or the administrators.

The user has also the ability to upgrade its account to premium. With a total cost of 290 GBP, the user sends a virtual request for an account upgrading. The Premium Account has more capabilities that are not analyzed further in this thesis. This option is just given for simulation purposes.

Another option of the user is to leave the grid and unsubscribe from the grid service. The user must check the checkbox in order to confirm that he/she wants to unsubscribe. Only if he/she checks the box, the “Leave Social AdGrid” button is able to be pressed. After that, a request is sent to the support, in order to remove the user from the grid service. This option is also provided for simulation purposes.

### 3.8 Grid Request – The Main Component

The option “Grid Request” that is only available to registered members of the grid service is the main function of our system. In short terms, it allows the user to search into the databases for users with specific criteria.

After the user logs in and selects the option “Grid Request” from the main menu, the Grid Service is initializing and the user will see a screen like the one in figure 3.12. The screen becomes darker and while the grid service is initializing, the user cannot have any interaction with the service’s interface, except the cancel button that is seen again in the figure. Once the grid service has initialized and is able to start running, the user will see a page like the one in figure 3.13.
Figure 3.12: Initializing Grid Service

Figure 3.13: Search Criteria
In this page, the user has to specify the search criteria. Data mining and pattern matching techniques will be applied to the virtual database, according to the criteria of this search.

The user has a variety of criteria to choose from, such as age, sex, hobbies, interests, country and education level. He can also choose in which social networks the search will be performed. As seen in figure 3.14, at this version of the simulator, there are three social networks that are the search can be performed: Facebook, LinkedIn and Twitter. The user has the ability to set the search to all three of them, but the cost of the search is increasing each time he adds one.

The most important feature of the simulator and the “Grid Request” component is that not only the user can search for people, but for Human Relationships as well. The simulator has been designed in order to simulate human relationship discovery. If the user checks the boxes of the figure 3.15, the grid service looks for people that are connected with any kind of relationships to the result individuals of the search. It is also possible to search for groups and teams that the people are members of. By clicking these checkboxes, the cost of the search is virtually increased.
3.9 Progress

After filling the criteria in the previous form and pressing the submit button, the user is redirected to the page that is shown in the figure 3.17. Let’s assume that the user wants to search people from Facebook and LinkedIn, with a choice of finding also the friends of the people. Other criteria are: a country of residence Greece, males of 24 years old with Ski as the main hobby. The criteria look like in figure 3.16.

In figure 3.17 we are simulating the procedures that take place after the user’s search request. It is assumed that the grid service is searching in two Grids, the “Social European Grid” and the “Social USA Grid”.

![Figure 3.16: User provided criteria](image1)

![Figure 3.17: Processing user’s request](image2)
Data Mining and Pattern Matching techniques are also being applied to the databases and their progress can be seen as well. Additionally, there is the total cost of this search, which virtually depends, as mentioned, in the time that the grid has spent in order to retrieve the results and in the criteria that the user has specified.

If the user wants to see the results, he/she needs to check the accept box. In that way, he/she accepts that the company’s bank account will be charged with the specific amount of money. After that, the “Show me the results…” button becomes available. Once the user presses it, the total charge and the total requests of his profile are changing. The total requests are increased by 1 and the total charge is increased by the amount of money that the user just accepted (in the figure 3.17 it is 83 GBP).

### 3.10 Search Results

After the user presses the “Show me the results…” button, the simulator is ready to show the results to the user. In figure 3.19, the page that prints the search results to the screen is presented.
From the figure 3.16, somebody can recall that the user had checked the “Find Friends” checkbox. That means that the grid service has searched for people that are related with the persons in the results. In simple terms, it shows the people that have any kind of human relationships with the result customers. As mentioned before, that is the main feature of the Social AdGrid grid service proposal.

The grid service has found eleven people with the specific criteria. From these people, only two have Ski as a hobby and are 24 years old and from Greece. This means, that the other 9 people are related somehow to these 2 people from the search results, with any kind of human relationship. We now present a small amount of search results, but if the grid service was really implemented, the results would probably be hundreds of customers.

But the functions of this grid service do not end here. As seen in the figure 3.19, the Social AdGrid user cannot see much of the customer’s personal information, but the name, surname, age and country. He/she is also capable of seeing the social network that every person has its account in. Also, in the column that is named email, the user can see a link named “Contact” and under the list, a button named “Contact all users”. These last two functions can lead to another functionality of the Social AdGrid grid service, the advertising component.

### 3.11 Advertisement

This feature has been initially designed in order to offer customer data privacy protection. Due to the unknown future behaviour of the user, the grid service cannot provide him/her with sensitive information, such as the email of the customer. This kind of information can be misused by the advertising agency and that would result a violation of the customer’s privacy.

By clicking to any of the two mentioned options, the user of the grid service is given the opportunity to advertise to: 1) the selected people (“Contact” link in email column) or to 2) the whole list of the people that have been found (“Contact all users” button, under the list). This will forward the user to the component of our system that is called “Advertisement”. In figure 3.20 a screenshot of this component can be seen.
As mentioned, the user can only advertise with this tool that is provided from the grid service. This form is completed by the user in order for to be processed by the system and sent as an email to the selected customers.

In this form, the user of the grid service must provide the company’s name, the company's email and the subject of the email. Also, there is an option for subscription to the company’s newsletter and an option to ensure that the sent email will not be marked as spam in the customer’s email inbox. This last option costs 20 additional pounds (GBP) and once checked, the money will be virtually retrieved from the company’s bank account automatically.
Furthermore, the user must specify the format of the email that is going to be sent. There are two supported formats. The first is plain text and the email is going to be sent as simple text to the customers. This is a safer solution because not every email client supports html. The second format is HTML and the user has much more choices for creating an email.

As mentioned, there is also the choice of offering a subscription to the customer that the user advertises to. There is a standard, no-cost subscription and a professional, paid subscription. The user of the service can choose no subscription, or either of those two.

The grid service provides a simple, but at the same time complete, text editor for the user to write the advertisement email, as seen in figure 3.21. Most of the main functions of an editor are supported such as bold-italic-underlined, font and background colour, links, images and tables. Also, several fonts and styles are available for use. A simple text editor is provided, because we do not want to focus in this particular part.

But the service offers one more feature, as seen in figure 3.22. The user can write his/her own HTML code for the email that is about to be sent if he/she checks the box that is right after the text area (“Want to write your own HTML?”). In that way, the user gains more freedom for the creation and editing of the advertisement email.
By pressing the button “Advertise” that is in the end of this page, the user agrees to send the email to the selected customers. After that, the email is virtually checked by the administrators of the service and sent to the selected people.

This was the last step of the described use-case of the simulator. The user of the grid service can see again the list of the customers of his search request and continue advertising. The total charge in his account information has changed as well as the total requests. He/she can now make a new Grid Request or he/she can log out of the system.

3.12 Use Case Diagram

In order to present those steps, they have been given in a use-case diagram below. The various actions that the user is able to perform can be seen in this diagram.

Figure 3.23: Use Case Diagram
### Use Case Name: Main User Actions

<table>
<thead>
<tr>
<th>Scenario: Make a grid request and advertise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triggering Event: User wants to advertise</td>
</tr>
<tr>
<td>Brief Description: Need to find users with specific criteria and advertise to them.</td>
</tr>
<tr>
<td>Actors: Advertising Agency (User)</td>
</tr>
<tr>
<td>Pre-conditions: User has a valid account</td>
</tr>
<tr>
<td>Post-conditions: None.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Flow of events:</th>
<th>Action:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step:</td>
<td>1. User logs in to the system by providing username and password.</td>
</tr>
<tr>
<td>2. User provides grid request.</td>
<td></td>
</tr>
<tr>
<td>3. User specifies the criteria.</td>
<td></td>
</tr>
<tr>
<td>4. User accepts terms and pays.</td>
<td></td>
</tr>
<tr>
<td>5. User views results and selects preferred people.</td>
<td></td>
</tr>
<tr>
<td>6. User advertises.</td>
<td></td>
</tr>
<tr>
<td>7. User Logs off.</td>
<td></td>
</tr>
</tbody>
</table>

| Exception Conditions: User does not exist in the system (access restriction). |
| Comments: User can also see his account, upgrade it or delete it. |

### 3.13 Main Use Scenario

In previous paragraphs we have seen the several actions that the user of the grid service can perform. In this paragraph those steps are seen in an order. The following figure presents graphically the mentioned actions for the search request case.

![Graphical representation of the Use Case](image)

Figure 3.24: Main user’s actions


3.14 Chapter Conclusion

The purpose of this chapter was to introduce the reader to the Social AdGrid grid service. We believe though, that in order to understand better the functionality of this grid service, the reader should have a hands-on experience with the simulator that is provided together with this thesis document. We also believe that at this version of the simulator (version 1.5) we are able to demonstrate to the reader the main functions and tools that the user of the proposed grid service can use.

As mentioned earlier, the reader of this manual-like chapter needs to keep in mind that all of these functions and tools are only for simulating purposes and all that is happening in the grid service website are managed by the “Social AdGrid Simulator v1.5”. The goal is to demonstrate how Social AdGrid would work, in order to present the thesis idea and this is why this simulator has been developed.
Chapter 4

System Development and Code Analysis

In this chapter, we present the most important parts of the simulator’s code. For obvious reasons the whole code of the grid service simulator cannot be presented, because it include hundreds of lines of code. We present parts of the code in php, javascript, as well as some html. This chapter is split into paragraphs and almost every paragraph describes a specific function of the system.

4.1 Programming the Simulator

The simulator has been programmed using HTML, PHP, CSS, javascript and jquery. Several tools have been used, such as Artisteer and Notepad++ for the template creation and the coding respectively. Furthermore, we used Xampp that gives us the ability of running Apache and MySQL as services on our system. For the management of the databases, phpMyAdmin has been used. The simulator is delivered in the version 1.5 after corrections, updates and bug fixes that have been done together by the author (programmer) and the supervisors of this Thesis. It works efficiently, without any errors and known bugs. It has been tested and works perfectly in 2 different browsers, the Mozilla Firefox (version 3.0.13) and Internet Explorer (version 8).

4.2 The Database

With the use of the tool phpMyAdmin, a database with two tables has been created. The name of the database is “grid_ws_users”. The names of the two tables are “users” (for the users of the grid service) and “social_users” (for the people that can be searched with specific criteria – “customers”).

Figure 4.1: Database and tables in phpMyAdmin
In the “users” table, for testing purposes, there is only the user “socialadv”, as seen in the figure 4.2.

![Figure 4.2: User “socialadv” in table “users”](image)

In the “social_users” table, there are much more users, in order to demonstrate the functionality of the grid service. Some of the table’s contents and fields can be seen in the figure 4.3.

![Figure 4.3: Users in table “social_users”](image)

We have given only some fields that would satisfy our simulation purposes, such as hobby, country, age, email, education etc. In the figure 4.4 we can also see the types of each field of the above table.

![Figure 4.4: Field types and collation](image)
The snippets that follow show the sql parts that create the tables of the database. The first creates the table “social_users” and the second the “users”.

```
CREATE TABLE IF NOT EXISTS `social_users` ( 
  `ID` int(15) NOT NULL, 
  `Name` varchar(20) CHARACTER SET utf8 NOT NULL, 
  `Surname` varchar(30) CHARACTER SET utf8 NOT NULL, 
  `Social_Network` varchar(30) CHARACTER SET utf8 NOT NULL, 
  `Age` int(3) NOT NULL, 
  `Country` varchar(30) CHARACTER SET utf8 NOT NULL, 
  `Hobby1` varchar(30) CHARACTER SET utf8 NOT NULL, 
  `Hobby2` varchar(30) CHARACTER SET utf8 NOT NULL, 
  `Interest1` varchar(30) CHARACTER SET utf8 NOT NULL, 
  `Interest2` varchar(30) CHARACTER SET utf8 NOT NULL, 
  `Sex` varchar(7) CHARACTER SET utf8 NOT NULL, 
  `Email` varchar(40) CHARACTER SET utf8 NOT NULL, 
  `Education` varchar(40) CHARACTER SET utf8 NOT NULL, 
  `Grp` varchar(11) CHARACTER SET utf8 NOT NULL, 
  PRIMARY KEY (`ID`), 
  FULLTEXT KEY `Hobby2` (`Hobby2`), 
  FULLTEXT KEY `Hobby1` (`Hobby1`) 
) ENGINE=MyISAM DEFAULT CHARSET=latin1;
```

```
CREATE TABLE IF NOT EXISTS `users` ( 
  `username` varchar(10) CHARACTER SET utf8 NOT NULL, 
  `password` varchar(10) CHARACTER SET utf8 NOT NULL, 
  `Email` varchar(50) CHARACTER SET utf8 NOT NULL, 
  `website` varchar(30) CHARACTER SET utf8 NOT NULL, 
  `company_name` varchar(15) CHARACTER SET utf8 NOT NULL, 
  `Total_Charge` int(20) NOT NULL, 
  `Total_Requests` int(20) NOT NULL, 
  `Account_Holder` varchar(50) CHARACTER SET utf8 NOT NULL, 
  `Bank` varchar(35) CHARACTER SET utf8 NOT NULL, 
  `AccountNo` varchar(50) CHARACTER SET utf8 NOT NULL, 
  PRIMARY KEY (`username`) 
) ENGINE=MyISAM DEFAULT CHARSET=latin1;
```

The above code parts are saved in the “grid_ws_users.slq” file. As it can be seen, they call the function CREATE TABLE. This function is accompanied by the name of the table, and the command “IF NOT EXISTS” which checks if the table exists in the database.
Also, the next two code pars show the functions that insert the records into these tables. The first snippet shows a part for inserting the records in the table named “social_users” and the second the records for the table “users”.

The INSERT INTO function followed by the name of the table, inserts values records with specific values in the specified table. Inside the brackets, the coder needs to specify also the fields of the table.
4.3 Log In and Log Out

In order to simulate the Log in and Log out functionality, the following code was used and is included in the file “logged_in.php”. When called, the following code is connecting to the database “grid_ws_users” and is searching in the table “users” the user that is trying to log in. Except from the following code, some more is included in the beginning of every .php file of the simulator.

```php
<?php
    session_start();
    $dbhost = "localhost"; //Name of Database Host
    $dbname = "grid_ws_users"; //Name of the database
    $dbuser = "root"; //Database User's username
    $dbpass = ""; //Database User's password

    //Connect to database
    mysql_connect($dbhost, $dbuser, $dbpass) or die("Not able to connect:", mysql_error());
    mysql_select_db($dbname) or die(mysql_error());

    if($_POST[action1]) //User pressed Login Button
    {
        //Selects the user with the specified username and password from the
        //database "grid_ws_users" and the table "users"
        $query = "SELECT * FROM users WHERE Username = ".".$_POST['username']."."
            AND Password = ".".$_POST['password'].".";
        $result = mysql_query($query); //Put the result into the variable $result

        if(mysql_num_rows($result) != 1) //If cannot find the user
        {
            $error = "Wrong username or password"; //Show error
            include "index.php";
        }
        else //If user has been found
        {
            //Put in the session variable "username" the value of the username
            $_SESSION['username'] = $_POST['username'];
            require "index.php";
        }
    } //end of if

    if($_POST[action2]) //If user wants to log out
    {
        //If user is logged in and the username is set
        if(isset($_SESSION['username']))
        {
            unset($_SESSION['username']); //Unset the variable username
            require "index.php";
        }
```
The following code is responsible basically for assigning the given username to a variable, named $un. Also, it gives to another variable named $myusername, the text “Logged in as” followed by the username. This code is used for the log-in form.

```php
<?php
    session_start();
    if (isset($_SESSION['username'])){ //If username is set
        $un = $_SESSION['username']; //assign it to the variable $un
        //and assign to the variable $myusername the text: "Logged in as + username"
        $myusername = "Logged in as ".$_SESSION['username'];
    }else{
        $un = "0"; //else assign 0 to the variable $un
        $myusername = "Not logged in"; //and "Not logged in" to variable $myusername
    }
?>
```

The code below is a mix of html and php code used for the log-in and log-out form in every .php file. If the user is not logged in ($un equal to zero), then the page shows the log-in form (see figure 3.8). If on the other hand, the user is logged in (which means $un other than zero), the log-out form appears (see figure 3.9).

```div class="BlockContent">
    <?php if($un=='0') { //if username is zero ?>
        <form action="logged_in.php" method="post">
            Username: <input type="text" name="username"><br>
            Password: <input type="password" name="password"><br>
            <input type="Submit" value="Log me in" name="action1">
            <br />
            Want to use the Service?
            <a href="apply.php">Apply Now!</a><br />
        </form>
    </div>
    <?php } //end of if($un=='0')

    else{
        echo $myusername; ??
    <br />
    <form action="logged_in.php" method="post">
        <input type="Submit" value="Log me out" name="action2">
    </form>
    <h4>You are an authorised user. You may use this grid service.</h4>
    </div>
    <?php } //end of else ?>
```

It is noticeable in the last snippet of code, that in each form, the proper action is followed (action 1 or 2). For example, if the user is not logged in and he/she presses the button “Log me in” then the ‘action 2’ is triggered from the logged_in.php file. With the same if statements we can control the access restriction to specific areas of the grid service (see figure 3.10).
4.4 My Account Data Retrieval

When the user is logged in, he/she can see the company’s account by clicking on the option “My Account” from the main menu. After that, the file account.php is responsible for retrieving the account from the database. This is accomplished with the code that is shown below.

```php
<?php
// Make a MySQL Connection
//Connection to localhost with username root and empty password
mysql_connect("localhost", "root", "") or die(mysql_error());

//Selection of database: "grid_ws_users"
mysql_select_db("grid_ws_users") or die(mysql_error());

$usn = $_SESSION['username']; //Assign to variable $usn the username

//Select from table "users" all of the users with username equal to $usn
//and assign it to the $result variable
$result = mysql_query("SELECT * FROM users
WHERE username='$usn'") or die(mysql_error());

//Assign to the variables the contents of the record
while($row = mysql_fetch_array($result))
{
    $accounter = $row['Account_Holder'];
    $companyname = $row['company_name'];
    $web = $row['website'];
    $email = $row['Email'];
    $reqs = $row['Total_Requests'];
    $charger = $row['Total_Charge'];
    $bank = $row['Bank'];
    $acnum = $row['AccountNo'];
}
?>
```

After the data is retrieved, the html shows them with the following code by using the echo php command:

```html
<h3>Account Details</h3>
<ul class="stats nobullet">
    <li><strong>Account Holder: </strong><?php echo $accounter ?></li>
    <li><strong>Company Name: </strong><?php echo $companyname ?></li>
    <li><strong>Website: </strong><?php echo $web ?></li>
    <li><strong>Email: </strong><?php echo $email ?></li>
    <br />
    <li><strong>Total Requests: </strong><?php echo $reqs ?></li>
    <li><strong>Total Charge: </strong><?php echo $charger ?><span> GBP</span></li>
    <br />
    <li><strong>Your Bank: </strong><?php echo $bank ?></li>
    <li><strong>Bank Account Number: </strong><?php echo $acnum ?></li>
</ul>
```
In the account.php file, there are more options besides the view of personal information (see paragraph 3.7), such as ‘update account to premium’ and ‘leave the service’ options. In the last option, the user needs to confirm by checking the checkbox, before pressing the “Leave Social AdGrid!” button (see figure 3.11). This cannot be accomplished by html or php code and javascript has been used in order to implement it. In the following image, it is shown the javascript function that has been used. With this script the pressing of the button is allowed, only if the checkbox is checked.

```javascript
<script type="text/javascript">
    var lst = ";"
    document.getElementById('textbox1').value = lst;

    function EnableDisableButton(check, un)
    {
        if (check.checked==1)
        {
            lst=lst+un+";"
            document.getElementById('textbox1').value=lst;
        }
        if (check.checked==0)
        {
            var vrb;
            vrb=";"+un+";"
            lst = lst.replace(vrb,";");
            document.getElementById('textbox1').value=lst;
        }
        if (lst==";"")
        {
            document.getElementById('submit').disabled=true;
        }
        else
        {
            document.getElementById('submit').disabled=false;
        }
    }
</script>
```
4.5 **Grid Request Search Criteria Form**

If the user is logged in, he/she then can select and use the “Grid Request” option from the main menu. If not, a message of access restriction appears. This is controlled partially by the following code.

```php
<?php
    if($un){ //if user is logged in ?>
        <span class="PostHeader">Place your request to the Grid... </span>
    } else { //if user is not logged in ?>
        <span class="PostHeader">Restricted Access!</span>
    </h2>
<?php } ?>
```

The form of the search criteria is made by HTML and there can be seen some parts of it in the following snippets of code.

```html
<form action="progress.php" method="post">
    <br />
    Social Network 1:
    <select name="sn1">
        <optgroup label="Social Network 1">
            <option value="">Not specified</option>
            <option value="Facebook">Facebook</option>
            <option value="LinkedIn">LinkedIn</option>
            <option value="Twitter">Twitter</option>
        </optgroup>
    </select>
    <br />

    Sex:
    <select name="Sex">
        <option value="">Not specified</option>
        <option value="Male">Male</option>
        <option value="Female">Female</option>
    </select>
    <br />

    Age:
    <input type="int" size=3 name="Age" />
    <br />

    Find Friends?
    <input type="checkbox" name="ch1" value="net"/>
    (Human Relationships Discovery)
    <br />

    Find Groups?
    <input type="checkbox" name="groups" value="Groups" />
    <br />
```
There are several criteria and selections that are given to the user, as seen in paragraph 3.8.

### 4.6 Search Request Processing Form

As seen in paragraph 3.9, in this form the system is simulating the “processing” of the user’s search request. In this form, there are progress bars in order to simulate the processing of searching in grids and of applying specific techniques. This is done by using an open source javascript, which name is jsProgressBarHandler.

In order to use this javascript, first the position must be declared in the beginning of the .php file and then use the progress bars with html commands, as we seen in the two code snippets below.
After the progress bars view, the cost of this specific search is generated (randomly) and added to the total cost of the user’s account. Also, the total requests that the user has performed must be updated by 1. These two actions are done by using php, as we can see in the following code.

```php
<?php
    // Make a MySQL Connection
    // Connection to localhost with username root and empty password
    mysql_connect("localhost", "root", 
""") or die(mysql_error());

    // Selection of database: "grid_ws_users"
    mysql_select_db("grid_ws_users") or die(mysql_error());

    // Assign a random integer to the variable, in range 50-140
    $charging = rand(50,140);

    // Assign the username to variable $usnm
    $usnm = $_SESSION['username'];

    // Update the field "Total_Charge" of the user that is logged in,
    // by adding the random integer that the function rand() returned
    mysql_query("UPDATE Users SET Total_Charge = Total_Charge + '$charging'
        WHERE username = '$usnm'";

    // Update the field "Total_Requests" of the user that is logged in, by 1
    mysql_query("UPDATE Users SET Total_Requests = Total_Requests + 1
        WHERE username = '$usnm'";

    echo $charging  // Print the cost
?>
```

The database searching is not actually taking place in this .php file (progress.php), but in the next, which is called (search_grid.php). From the previous php snippets, it has been shown that in the search criteria form (see figure 3.13) after the pressing of the submit button, the user provided the system with specific search criteria. The data from that .php file (grid_request.php) passed in this php file (progress.php) with the post method that is shown in this line of html code:

```html
<form action="progress.php" method="post">
```
The problem here, is that there can be retrieved only the data in the current .php file (progress.php) and cannot retrieve them in the following one (search_grid.php) with the post method. Another technique needed to be used in order to pass the data to the next .php file. The method that is used is saving the search criteria data to the Session table (or Session Array). In php, a session array is a method of maintaining data in the memory and retrieving them across the .php files. Graphically, the data passing in our system is like the one in the figure 4.5.

![Figure 4.5: Field types and collation](image)

The Session table method can be seen in the following code snippet. Also, the control methods for checking if the user wanted human relationships searching, can also be seen.

```php
<?php

//Take all the variables from the form in grid_request.php
//and add them to the SESSION table for use in search_grid.php
$_SESSION["hobby1"] = $_POST["hobby1"];
$_SESSION["hobby2"] = $_POST["hobby2"];
$_SESSION["interest1"] = $_POST["interest1"];
$_SESSION["interest2"] = $_POST["interest2"];
$_SESSION["country"] = $_POST["country"];
$_SESSION["Sex"] = $_POST["Sex"];
$_SESSION["Age"] = $_POST["Age"];
$_SESSION["sn1"] = $_POST["sn1"];
$_SESSION["sn2"] = $_POST["sn2"];
$_SESSION["sn3"] = $_POST["sn3"];
$_SESSION["education"] = $_POST["education"];

//The following code is to check of the user has checked
//the "Find Friends (Human Relationships)" checkbox
if (isset($_POST["ch1"])) {
    $ch1 = $_POST["ch1"];  
    if ($ch1 == 'net') {  
        $ch2 = 'checked';  
        $_SESSION["ch2"] = $ch2;
    }
} else {
    $ch2 = 'unchecked';
    $_SESSION["ch2"] = $ch2;
}
?>
```
4.7 Mining the Database

In this part of the system, the database is mined for finding records with the specific criteria that the user provided in the search form. There are mainly two cases that are simulated in this system. Before present the two cases, the variables need to be taken from the Session table as mentioned before. They are stored in variables with specific names and a connection with the database is established.

```php
//Assign to the following variables the data from Session table
//that were stored in progress.php file.
$hobby_1 = $_SESSION["hobby1"];  
$hobby_2 = $_SESSION["hobby2"];  
$interest_1 = $_SESSION["interest1"];  
$interest_2 = $_SESSION["interest2"];  
$country = $_SESSION["country"];  
$sex = $_SESSION["Sex"];  
$age = $_SESSION["Age"];  
$sn1 = $_SESSION["sn1"];  
$sn2 = $_SESSION["sn2"];  
$sn3 = $_SESSION["sn3"];  
$education = $_SESSION["education"];  

//Also assign the control of human relationship discovery
$friends = $_SESSION["ch2"];

// Make a MySQL Connection
//Connection to localhost with username root and empty password
mysql_connect("localhost", "root", ") or die(mysql_error());

//Selection of database: "grid_ws_users"
mysql_select_db("grid_ws_users") or die(mysql_error());

// Get results from the DB
if ($friends == 'checked') //If "Find Friends" is checked
{
    $result = mysql_query("SELECT * FROM social_users
WHERE
Sex='"$sex"
AND Country='"$country"
AND Age='"$age"
AND Hobby1='"$hobby_1"
AND Social_Network='"$sn1"
AND Grp = 1
") or die(mysql_error());
}
```

The first case is the one that the user has selected the search of human relationships and is seen in the following part.
In the second case, the user has not selected the finding of human relationships and the system searches in the database only for people with the specific criteria.

```php
// If the option is not checked then find only people with specific criteria
else
{
    // Select people from table "social_users"
    $result = mysql_query("SELECT * FROM social_users
    WHERE
    Sex='\$sex'
    AND Country='\$country'
    AND Age='\$age'
    AND Hobby1='\$hobby_1'
    AND Social_Network='\$sn1'
    OR Social_Network='\$sn2'
    OR Social_Network='\$sn3'
    AND Hobby2='\$hobby_2'
    AND Interest1='\$interest_1'
    AND Interest2='\$interest_2'
    ") or die(mysql_error());
}
```

### 4.8 Showing Search Results

After the data retrieval from the database, the system presents the results to the user of the grid service. This is accomplished with the following code snippet.

```html
<table border='5'>
    <tr>
        <th>Name</th>
        <th>Surname</th>
        <th>Social Network</th>
        <th>Age</th>
        <th>Country</th>
        <th>Email</th>
    </tr>

<?php
    // Here we are printing the results inside the table
    while($row= mysql_fetch_array($result)) {
        <tr>
            <td><?php echo $row['Name']; // User's name ?></td>
            <td><?php echo $row['Surname']; // User's surname ?></td>
            <td><?php echo $row['Social_Network']; // User's SN ?></td>
            <td><?php echo $row['Age']; // User's age ?></td>
            <td><?php echo $row['Country']; // User's country ?></td>
            <td><?php // Email not visible or privacy reasons
                // instead, contact user option available ?
                <a href="send_email.php">Contact User</a></td>
        </tr>
    }
    </tr>
</table>
```
First a table is created with html in order for the results to be put inside. Then the results are printed with the ‘echo’ php function in every column of the table. Instead of showing the email (privacy reasons, see paragraph 3.10 and figure 3.19) a link is provided with the name “Contact User” in every line. Also, a button is provided too, for contacting all users at once:

These links lead to another page that is examined in the next paragraph.

4.9 Advertising Email Form

In this part of the system, the user is given the ability to advertise to the selected users. The user is able to see a form and a text editor with several functions. The editor is an open source javascript called “Cross-Browser Rich Text Editor” that we edited to fit our purposes. In the code below it can be seen the form and the calling of the script to make the editor. First of all, the position of the script is declared.

Then the form for advertising is created, as seen in the code that follows. In this form it is also called the script for the creation of the rich text editor.
After the script, the rest of the advertisement form is continued (see paragraph 3.11 and figure 3.20). The following code is printing in the screen the selections and choices that the user must fill in, in order to send the advertisement email.

The code “writeRichText(...);” is a function that calls the creation of the rich text editor with the specifications inside the brackets.

```html
<form>
  Your Company's Name: <input type="text" name="Name" value="" size="25" maxlength="50" />
  <br>
  Your Company's Email: <input type="text" name="Email" value="" size="25" maxlength="50" />
  <br>
  Email Subject: <b> <input type="text" name="Email" value="" size="50" maxlength="80" </b> <br>
  <br>
  <input type="CHECKBOX" name="subscribe"> Ask user to subscribe to your newsletter. <br>
  <br>
  <input type="CHECKBOX" name="subscribe"> Ensure not to be marked as Spam. (+20 GBP if checked.) <br>
  <br>
  Choose the format of your Email: <br>
  <input type="radio" name="format" value="html" checked> HTML <br>
  <input type="radio" name="format" value="text" checked> Plain Text <br>
  <br>
  Type of subscription you want to offer the Client: <br>
  <select name="type">
    <option value="not"> No Subscription </option>
    <option value="standard"> Standard - FREE </option>
    <option value="prof"> Professional - Paid </option>
  </select>
  <br>
  <br>
  Your Advertisement Here</b>
  <br>
  <br>
  <script language="JavaScript" type="text/javascript">
    writeRichText('rtel', 'place your "<em>ad</em>" here', 100, 150, true, false);
  </script>
  <br>
  <input type="SUBMIT" name="submit" value="Advertise!!">
</form>
```
4.10 Apply to Join the Grid Service

As seen in paragraph 3.3, it is given to the non-registered users the ability to join the grid service. This takes place in apply.php file and in this paragraph the most important parts of this code are presented. In the following snippets, there is the form that the user must fill in, in order to apply for joining the service (see figure 3.4). In this first part, the form is declared and begins. The name of the form is “applyform”.

In the next two parts, the Account holder, the company details and the address are presented in the screen in order to be filled by the user.
In the following code part, the applicant declares the company’s status. Some choices are given by default.

Also, the user is requested to fill in a text with reasons for joining the Grid Service.
4.11 Contact Email Retrieval

In the contact form, if the user is logged in, his/her email is automatically retrieved. If the user is not logged in, the email field is editable and the user must provide an email to contact the support (see figure 3.5 and 3.6). This is accomplished with the code parts below.

If the user is logged in, a connection with the database is established and the simulator searches in the database for the user that is logged in. When found, the user’s email is retrieved and shown in the screen.

```php
<?php
if($un)
{
    $email_show = $SESSION['username']; //Assign username to variable

    // Make a MySQL Connection
    //Connection to localhost with username root and empty password
    mysql_connect("localhost", "root", "]") or die(mysql_error());

    //Selection of database: "grid_ws_users"
    mysql_select_db("grid_ws_users") or die(mysql_error());

    //Select from the table "users" the records with username the
    //one as the variable's $email_show
    $result = mysql_query("SELECT * FROM users
WHERE username='$email_show'") or die(mysql_error());

    while($row = mysql_fetch_array($result))
    {
        echo "Your email is: ";
        echo $row['Email']; //Print email
    }
}
else //if user is not logged in
{
    <label class="desc" id="title2" for="email1">Your Email Address</label>
    <input id="email1" name="email11" type="text" value=""
    maxlength="255" tabindex="2"/>
</div>
<?php } ?>
```

If the user is not logged in, simply the email address textbox appears.
4.12 Loading Panel Script

In order to simulate the initialization of the grid service, it was needed to provide a loading progress bar, when the user pressed the “Grid Request” option from the main menu (see figure 3.12). To do that, we used an open source javascript called YUI Loading Panel. The panel is called from the .php files, with the following code, inside the script keywords.

```
<li><a href="index.php">Home</a></li>
<script type="text/javascript">
    var loadingPanel = new yuiLoadingPanel();
    loadingPanel.cancelEvent.subscribe(function(e, a, o){
    });
</script>

<li><a href="grid_request.php" onclick="loadingPanel.show()">Start Grid Service</a></li>
<li><a href="company.php">Contact</a></li>
```

As seen in the above snippet, this is the top menu bar with 3 options: Home, Start Grid Service and Contact. The same script can be called from the main menu as well.
4.13 Chapter Conclusion

In this chapter, we presented the most important parts of the code that was programmed or used in order to develop the “Social AdGrid Simulator v1.5”, which simulates the idea that is proposed in this Thesis. Most of the code of this system is written by hand. Also, open source javascripts and parts of code, were used in order to achieve some of the programming goals, such as progress bars and word editors, but this is a very small part of the system. Regarding these parts, the code of the open source scripts had to be edited, in order for them to fit our needs. More information for these parts is given in the Appendix B. The purpose of this chapter was to give a clear view of the system’s code, in order for the reader to understand how the various components of the simulator work.
Conclusions and Future Work

This thesis proposes a grid service that combines mainly four science sectors: Grid Computing, Social Networks, Human Relationships and Targeted Advertising. The service proposal is using Grid Computing in order to apply a variety of very demanding techniques in social network databases. In this way, human relationships can be discovered in these databases and targeted advertising methods can be performed. Most of the topics that are related with the idea’s research and development process have been covered in previous chapters. These topics can be grouped in the following parts:

- Overview of the technologies that have been used for our proposal.
- Simplified description of the grid service idea.
- Grid service simulator manual-like demonstration.
- Grid service simulator’s code analysis.

Each of these parts has been described in detail and it has been illustrated with figures, screenshots and examples when necessary.

Conclusions

The goal of this thesis project was to propose a new idea that would use Grid Computing Technology in the field of Social Networking. The idea was simple in the beginning, but it became more complicated as it went on. After coming up with the idea, it had to be shown how this would work in reality and a simulator was developed in order to demonstrate this.

The simulator general objectives have been achieved. It has been managed to provide a very simple but at the same time efficient user interface for discovering people and human relationships with user-specified criteria. The users of the service have access simply though their web browser, no matter the brand and the version of the browser. Also, the simulator application has been designed with the ability to accept easily further changes or additions of new code parts, scripts or modules.
The solution provides a very new combination of technologies and gives a new vision in online targeted advertising. We believe that such a project could possibly attract the future attention of advertising agencies, which would like to optimize their ability of finding target groups and discovering human relationships. We also believe that data mining and human relationship discovery in social networks would definitely play a key role in the future of computer science.

During the research and development process of this Thesis Project a lot of knowledge has been gained in sectors like Grid Computing, Social Networks, Data Mining and Marketing. Our understanding of the term Grid Service and Social Network has also been improved significantly. Furthermore, we have been given the opportunity to research on the field that examines Human Relationships on the web.

Web designing skills have been developed in order to program the web-based simulator and languages like html, php, sql and javascript have been used for this purpose. Moreover, it has been managed to expand the borders of computer science and to possess knowledge in the marketing field and more specifically in targeted advertising methods and techniques. It is believed that the goals of the Thesis Project have been met and the proposed idea was presented in the best possible way.

In our opinion, this thesis proposal is able to find future applications in various domains of internet advertising and marketing, because it combines provides a very useful and efficient human relationship discovery and advertising platform.
**Future Work**

There are a lot of things that can be done in order to expand this thesis idea. First of all, it has been presented the first version of the idea and it has been given to the user of the grid service the ability to advertise only through email. The advertising part could be done directly in the customer’s social profile with profile embedded ads or other ways.

At this stage, only people that are generally related with the persons that fit our criteria can be found. A very efficient approach would be that the user of the service could specify the type of the relationship that wants to be found, such as family, friends, colleagues.

Furthermore, in the simulator there can be a larger number of criteria for customer detection. The larger the number of criteria, the more accurate the searching procedure and of course the more accurate the following advertising process.

Probably the greatest challenge is to avoid spam-like behaviour from the advertising agencies that would use this service. Today, there can be seen many companies and advertising agencies that are taking advantage of spamming. In this proposal we tried to give a solution to this problem, as described in previous chapters. But it is strongly believed that further solutions could be given.

The collected information at this stage of the idea is mainly profile personal text data. This can be expanded in photo and video data collection, in order later to apply identification techniques to discover related people and user’s preferences, by using face and object recognition. This could also be provided by the grid infrastructure as another grid service that would be applied in the database of the proposed grid service.

Closing, the proposed grid service could be a part of a greater group of social-grid-services that are offered by a company or an organization in a Grid.
References


Appendix A

Environment Configuration

This appendix briefly describes the installation and configuration procedures of the software that has been used in order to develop the project and we also mention the software that we used.

A.1 Apache and MySQL installation

To develop this project, Apache HTTP server and MySQL database have been used. In order to install and use these services, it has been used an open source cross-platform web server package called XAMPP (X-cross platform Apache HTTP Server MySQL PHP Perl). Xampp is released under the terms of GNU General Public license. It can be found at http://www.apachefriends.org/en/xampp.html.

A.2 PhpMyAdmin

For the management of the databases a tool called phpMyAdmin has been used. This tool is a free software that is written in PHP. Its purpose is to handle the administration of a variety of operation with MySQL, such as database management, tables and fields, indexes, permissions and users. The version that was used is the 1.7.1 for Windows and it can be found at http://www.phpmyadmin.net.

A.3 Notepad++

For the writing of the code, we have used Notepad++. This is a source code editor that is written in C++ and runs under Microsoft Windows. It is distributed as free software. This tool has been used in order to write and edit php, html, sql and javascript code. It can be found at http://notepad-plus.sourceforge.net/uk/site.htm.
Appendix B

Scripts and Widgets

This appendix shows the scripts and widgets that have been used in the development of the simulator. The scripts can be found on the internet and they are free to use.

B.1 jsProgressBarHandler

In order to simulate the grid searching progress, this javascript has provided the simulator with the proper progress bars. It is free to use and can be found in: http://www.bram.us/projects/js_bramus/jsprogressbarhandler

B.2 Cross-Browser Rich Text Editor

In order to embed a text editor for the advertisement component, this javascript provides a simple but very useful rich text and html editor. It has been edited by the programmer in order to fit the simulator’s needs. It is free to use and can be found in: http://www.kevinroth.com/rte/

B.3 YUI Loading Panel

For the simulation of the grid service initialization, this widget provided a progress bar loading panel. It is free to use and can be found in: http://thecodecentral.com/2008/01/17/a-yui-loading-panel-widget