Lappeenranta University of Technology

School of Engineering Science

Master's Programme in Computer Science

Jai Kumar

Matchmaking Database for CST

Examiners: Professor. Ajantha Dahanayake
               Samantha Kiljunen,PhD

ABSTRACT

Lappeenranta University of Technology

School of Engineering Science

Master's Programme in Software Engineering and Digital Transformation

Jai Kumar

Matchmaking Database for CST

Master's Thesis

65 pages, 31 figures, 2 tables, Reference 48

Examiners: Professor. Ajantha Dahanayake
                Samantha Kiljunen, PhD

Abstract:

Center of Separation Technology (CST) Department realized the need of Matchmaking application. The main purpose of application was to fulfill the needs of members, users and concerned people within this industry. The need was to finding a platform where experts, industry people, new users in the industry, etc., are gathered at same platform and can easily find each other. The application is developed and ready to use for all the concerned users and if users have very small knowledge of what exactly they are looking for, then they can use this database and find the solutions. This implementation is an initial version of this application; As future developments there are, various options are possible for the application such as, combining with artificial intelligence that will help in analysis of the behavior of user search patterns and making future recommendation about experts, relevant industries.

ACKNOWLEDGEMENTS

Table of Contents

List of Figures

List of Contents

LIST OF SYMBOLS AND ABBREVIATIONS

ACL   Access Control List

BSON   Binary-Encoded Serialization of JSON

DR   Deputy Registrar of the Supreme Court or the Subordinate Courts of Singapore

CST   Centre of Separation Technology

DVI   Direct visual inspection

EXIF   Exchangeable Image File Format

HTML   Hypertext Mark-up Language

HTTP   Hypertext Transfer Protocol

LUT   Lappeenranta University of Technology

NEET   Not currently Engaged in Employment, Education, or Training

NHL   National Hockey League

RAD   Rapid Application Development

RDF   Resource Description Framework

RQ   Research Question

SaaS   Software as a service

SWSE   Semantic web search engine

URI   Uniform Resources Identifiers

USB   Universal Serial Bus

WWW   World Wide Web

W3C   World Wide Web Consortium

XML   Extensible Markup Language

1 INTRODUCTION

The main aim of the thesis is development of a Matchmaking application and information database. The initiative for creating application and database taken by the Center of separation technology (CST) Lappeenranta University of technology (LUT). The Importance of Matchmaking application and centralized database realized by CST. The purpose of application is easy access of information for users connected with chemical and separation industry.

The objective of the work is to design a web based user interface with emphasis on web application for accessing the matchmaking data. Other objectives of the thesis are to study technical background of web-based tools, data, database tools and cloud tools such as Google cloud, Microsoft Azure and Amazon web services (AWS). It gives detailed knowledge of technical features of practical side of the project work..

Two important means of interaction in traditional hypertext web are searching and browsing. While browsers run the mechanisms for routing the information space, the place for the beginning of navigation process are often search engines. Many search engines have developed that crawl connected-data from web by providing query abilities over accumulated data and following resource description framework (RDF) links. There can be two categories of these services, Application oriented indexes and human oriented search engines [1].

Human oriented search engine is an engine that require human interaction to filter the results requested by humans and support user in clarifying search request made by humans [2]. Keyword based search services provided by search engines such as Semantic web search engine (SWSE) and Flacon towards human users and current market leaders like, Yahoo and Google follow same interaction paradigm. Search box are provided to user so that they can enter keywords they interested or any topic for items they are looking for. Falcon and SWSE not only just simply provide search result links but they also provide detailed interfaces that exploit the core structure of the data [1]. It helps the user in getting better insight into the information they are looking for.

Search proficiencies provided towards human by Falcon and semantic web search engine (SWSE) are alternative breed of resources that have been designed/developed to serve the requirements of applications built on top of distributed Linked Data. Swoogle and Sindice are the application-oriented-indexes that provide Application program interface (APIs) through which Interrelated Data applications can determine RDF documents on the internet that reference a certain URI or contain certain keywords. The foundation of such service is that it does not require linked data application to implement its own foundation or base for indexing and creeping all part of web of data to make it use [1].

The matchmaking application for CST is relevant to a Human oriented search engine. The users get two different options, first is selection choices from a drop down menu and the other is a search box filtration. The user can search the data using different key words and also they can filter out from the prior search result accordingly. The match making application created using different technologies such as, Python, HTML, CSS, JQUERY, JavaScript, Ajax, Microsoft Azure and database Mongo dB for storing data.

The most common way of implementing service-oriented concepts are web service technologies. Many researchers have recommended the usage of semantic information or data in semantic web service, resulting in the conception of semantic web services (SWS). From semantic web, it is reasonable to apply method, techniques and tools, it is an extension of web 3.0 (Current web) as proposed by Berners-Lee [3]. The Logical semantic web activities of the w3c is to enhance information on the internet/web with a distinct meaning and a layer of machine readable data [3].

Service discovery is one of the prime applications of SWS, it is largely affected by three steps.
   i.   Service provider has the ability to describe their services.
   ii.   Requester have the ability to describe their conditions towards services.
   iii. Efficiency of the service matchmaker.
An example of above mentioned steps:

An Algorithm takes a request from a user interface using a keyword and considering it into account, and finds the most suitable service as result from set of offered services [3].

The basic concept of matchmaking is matching two or more different entities. It is the process of finding to fulfill the required purpose of users. The purpose varies for different industries such as, Web service and application, Sports, boxing, Online games, medical for pairing organ donors. Many researches show, matching is a method of fulfilling the service request received from different entities or keywords. Such a process is also said to be a matching engine or matching tool. Concerning this thesis or semantic web matchmaking application or services is the process of finding the most suitable experts, process of filtration and helping individuals in finding matching projects according to interest and requirements. The experts can also look for the projects and thr candidate's profile for them to work together and fulfill each other's requirements. While developing this application agile methodology was followed for development. The application developed was presented in the CST event. More detailed procedures and feedback are discussed in chapter 4 of this thesis.

1.1 Motivating Example
To elaborate the idea of a web application and Matchmaking database for CST industry, it describes a motivating scenario and example. The application is developed for the chemical department especially for "Separation industry" which is itself a vast field of study. If anyone requires information regarding separation processes or experts, it is time consuming. Most of the time trying to find the information repeatedly on google gives us the same result, which does not produce the outcomes.

Case1:
A company wants to remove chemicals from water and wants to make some research on the separation process. As the company is not sure about the experts, when they start searching they get many results on search engines, but on the company websites there is too much information and not clear which experts is

specialized in which industry? It becomes difficult for the companies to find out about experts, if they keep searching repeatedly, they get the same results and content.

Case 2:

Some companies know about the industry, separation process, separation target or both of them. Still it is not an easy task for them to find out the exact possible solution for their problem for separation process. They choose any of the options and at the end figure out the problem is not completely resolved or properly researched according to the needs of the company.

1.2 Research question

- How to create an application that matches the experts?
- What can be the possible solution be to find an expert if industry, separation target or separation technology is known?
- How matchmaking application can be helpful for the industry, individual or organizations?

2 BACKGROUND

The technology has evolved very rapidly in recent times; previously it was just a vision to work on advanced technologies, but now it is a reality. Storing and managing data is one of the biggest achievements of today's era, in the past many innovative data storing products were used such as Punch cards, magnetic tape, magnetic-drums and floppy disks. In 1956, IBM introduced the first data storage device for the computer, after that it marked rapid progress and it was a kick-start for data storage industry [4], [5].

Using the data storage devices capabilities and adding different layers of hard drive device (HDD) storage systems created. In 1956 when these HDD were built, it was not popular due to the fact they were large and expensive. In the late 1980s, the prices and size started to reduce day-by-day [6]. The internet was designed in the 1960s, and during 1990s, the invention of the World Wide Web (WWW) took place and making available the feature of a virtual backup service, which meant devices were not required to be carried anymore and data could be accessed and backed up via remote connection.

2.1 Databases

During 1960s, first appeared in research papers, with the term "Database". Over a decade, dominance, surveillance and regulation has been symbol of large institutional databases in U.S . Deployment from military to analysis of credit score of individuals, database is supposed to serve as a dehumanizing role, converting the individual focus into a controlled set of statistics. During 1970s and 1980s, database was also significant technology in the conventional concept of personal computing created by microcomputer researchers, fans, entrepreneurs and hobbyists, mainly the big institutions were initial users of database technology. In 1990s, the database design was reverted from desktop, the promises made with populist were completely overlooked and database technology completely became institutional technology [7].

Computer researchers and scientists distinguish between database and database management systems (DBMS) very carefully. Database, which strictly deals with data storage whereas DBMS deals with managing of data and it also possess tools for

evaluating, organizing and controlling the data in that system. In order to simplify the historical comparison, generally database can be defined as managing data according to a standard model and loading it in a material form. Considering systems before 1960s, databases were fictional, but it highlight the endurance among related terms such as Data management and information processing [7].

Whenever we hear about any entity, we retrieve it using image in our brain it has been revealed by many researchers. For example, if someone ask you, do you like Grapes? You will think information and recall an image of grapes in your mind, instead of that, you will not start thinking about the alphabets "G", "R", "A", "P", "E", "S" [8]. Most probably, the data processes in our minds works the same way, obviously it is complex but highly effective at the same time. Now what would be your answer if I ask you to define "database"? The most obvious positive answer that comes tomind would be collection of tables which are relevant to each other. If this question had been asked from someone back in 1990s, probably their thoughts would be big single table, where they can retrieve all the information. Now the role of database had been changed in the society of current era. The database and applications have been changed a lot over the time [9].

2.1.1 Evolution of Databases
Database system have grown to set an oriented system from record-aligned directional database systems, hierarchical and network systems that provide path to relational database system. The relational database systems are now evolving into multimedia and object relational database system.
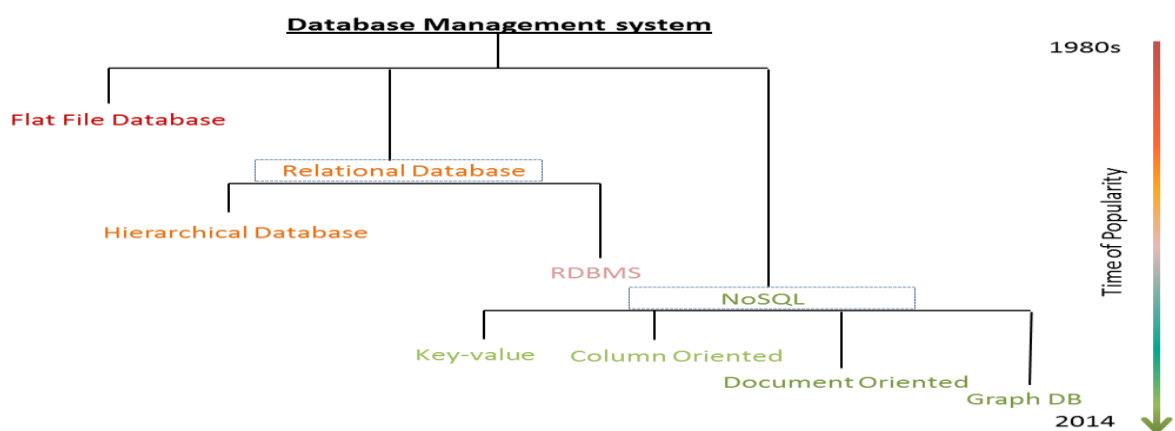

Figure 1. Evolution Tree of DBMS (Michael Waldman 2015)

The above tree will help us in understanding more about the evolution and mapping of all types of system for database management in a timeline [9].

2.1.2 Relational databases

Soon people started understanding that the flat file system cannot be used r long term. The flat file system was creating redundancy in data at every entry point. People considered a different way for data storing in different tables and defined an hierarchy to retrieve the data and it was called an hierarchical database. Hierarchical database and resembles a folder structure in laptops and systems. Each folder can contain a sub folder and a sub- folder has more sub folders and finally an hierarchy of dataset is created.
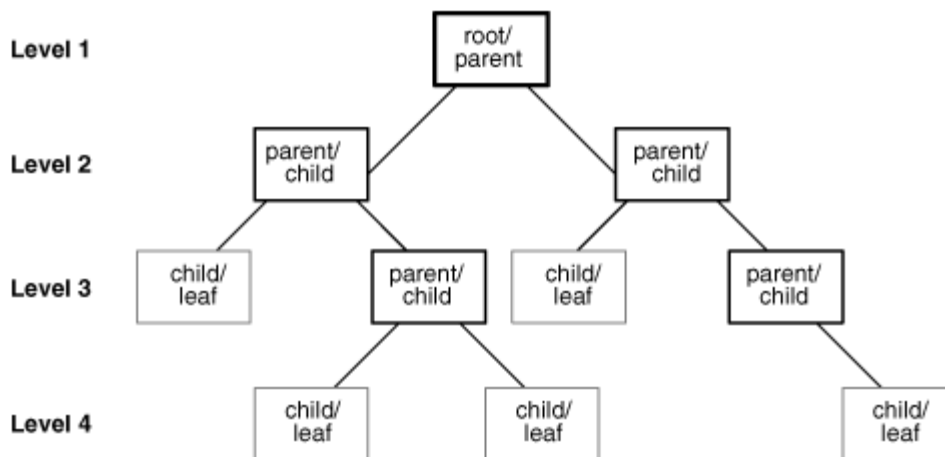

Figure 2. Hierarchy database structure (Tavish, Srivastava 2014)

An Hierarchical database is helpful in solving many problems. Its usage is limited to a one to one mapping data structure [10]. In addition, this mapping fails if more than one relation is required, therefore people thought about a database, which can have various kinds of relations. Such a structure is possible with one to many mapping.

Relational Database Management System (RDBMS) is known for these kind of tables and data structures [8]. The term relational database was introduced in 1970's by Edgar Codd's in his Research paper "A Relational Model of Data for Large Shared Data banks" Following figure shows two tables in simple database, one has department related records the other has employee related records. As its shown in Fgure 3 the link(join) is made by Department id between two tables, representing primary key in Department table whereas shown as foreign key in employee table. It

also illustrates a one to many relation in department tables, which shows employee tables have many records. The data in RDBMS is maintained by Structure query language (SQL). Example of SQL is as follows: **SELECT ∗ FROM employee WHERE department_ID = '22';**

Employee ID	Surname	Given Name	Initial	Department ID	DOB	Salary
35411	Felsky	Warren	J	22	25/2/1983	21000.00
35567	Cash	Martin	B	21	1/6/1989	19750.00
35582	Kelly	Edward	T	22	18/7/1961	45450.00

Department ID	Department Name	Department Location
21	Accounts	Head office
22	Payroll	Head office
23	Sales	Latrobe

Figure 3. RDBMS Example (Michael Waldman 2015)

2.1.3   NoSQL Database

Not only SQL is known as NoSQL. When it was realized by the people that unstructured data is carrying lots of data, which is difficult to mine via RDBMS. So people star finding ways of maintaining such datasets. Nowadays everything that is not linked with RDBMS is roughly known as NoSQL. NoSQL is known as the database of this current era, it was previously discussed briefly in 1998 in research papers andcompletely introduced in 2009. It helps in managing data, provided from current generation web services, such as Social Media, Blogs, etc. Below is the example where it shows why it is quite difficult to store information on RDBMS and it is example of social media, which has gained importance these days [8].
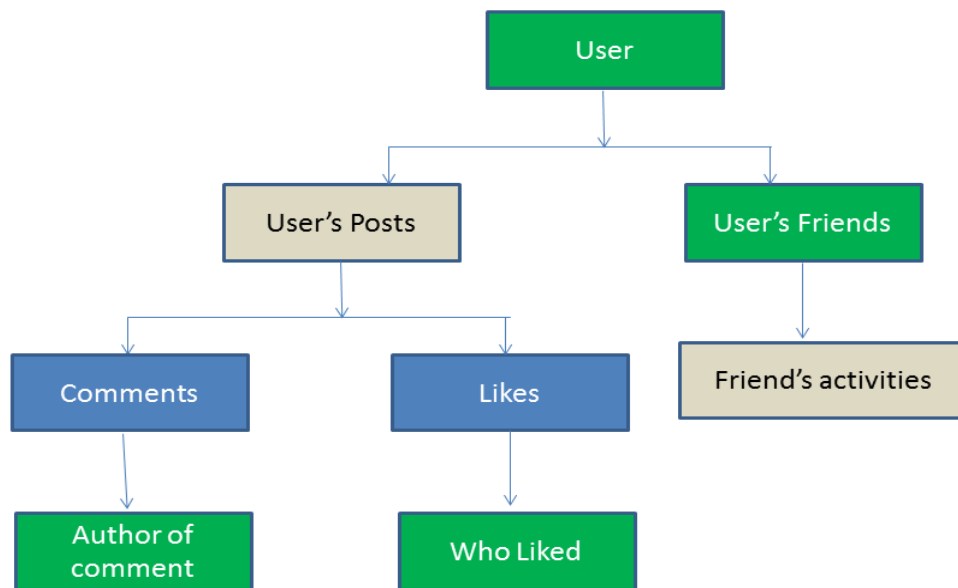
Figure 4. Usage example of NoSQL in social media [8]

The above figure shows the example of Facebook where one friend makes one post that others like and comment, the same color boxes represents the same category. If we wanted to query such kind of relation in RDBMS we need to assign many joins and with billions or trillions of rows just in order to get the home or activity page of user. For such kind of table we need to use more graph or flow based data structure. NoSQL structure has that benefit in it.

2.2 Web services

Web services are one of the fastest growing key blocks in business integration. Various industries and individuals develop enhanced and better web services using their own and others web services, which are accessible and reusable via the internet or intranet. These services are easy available on various networks.[11]
Every user has a different viewpoint regarding web technology and web services nowadays. Tim Burners-Lee introduced the web in late 1989, and he expressed his innovation in three interconnected phases: Web 1.0, Web 2.0, Web 3.0, named as Web of documents, Web of people and Web of data (the still-to-be-realized Web 3.0) respectively . The web is evolving day-by-day and is getting more data centric with the context of web 3.0. [12]

As mentioned above web 1.0, 2.0 and 3.0 are inter connected with each other. Web 1.0 is known as the initial footstep in web and possesses some qualities of Web 2.0, which has some additional elements of web 3.0.

2.2.1 Web 1.0:
The first generation of web introduced in 1989 as Web 1.0 lasted until 2005. The first generation of web www is defined as:

"An information space in which the items of interest referred to as resources are identified by global identifier called as Uniform Resources Identifiers (URIs) "

The purpose of the first generation was to deliver static webpages and content distribution. The main concept of web 1.0 was to find the content and read it. In other words, web 1.0 known as web information connections.

Hypertext transfer protocol (HTTP), Hypertext markup language (HTML) and URI are the core web protocols of web 1.0. The major characteristics are as follows:

- Read only content
- Creates an online availability of content that is available anywhere, anytime and to anyone.
- It uses simple Hypertext markup language and static pages

Web 1.0 has some limitations, discussed as follows:

- Only humans can read the pages in web 1.0, as the content was not machine compatible. Machines were not able to read the content in the pages of web 1.0.
- The web developer is only liable for managing and updating users and content of web pages.
- Dynamic demonstrations were not available to perform any dynamic events. No console support was present either.[12]

2.2.2 Web 2.0:
According to Dale Dougherty, the Web 2.0 was a read-write web as per his definition in 2004 [12]. Web 2.0 is not only the second generation but also a newer version of Web 1.0 that brought about drastic changes for the computer industry as it manages universal crowds in social communication with identical interest. Tim O'Reilly mentions in his website [13]:

"Web 2.0 is the business revolution in the computer industry caused by the move to the internet as platform, and an attempt to understand the rules for success on that new platform. Major rules are Build applications that harness network effects to get better the more people use them."[12]

It has all the major qualities like social, digital, relationship technologies and sharing which, in other terms, defines it as wisdom web. Web 2.0 has bi-directional transactions, which means it has the capability of reading and writing on the web. It has the ability for users to publish, manage and analyze content. The user can review the content provided by others.[14]

Web 2.0 can be categorized under various definitions in order to get a more complete understanding.

- **Technology-Centric:**

  It has become a platform not just for devices, it is way more than that. Some of the Technologies associated with it are Podcasts, wikis, blogs, etc. It focus on usability with user-generated content, not just by experts but also by anyone.

- **Business-Centric:**

  The internet as a platform is a big revolution for businesses in the computer industry. The user, web application and content promote social connection between individuals and organizations.

- **User-Centric:**

  Social websites and application consist of different communities; users can communicate, discuss, analyze, etc. It is the fastest way to share information and interact with people around the world. Web applications assist people in growing their social network and interchanging information.

It has been discoveredin different researches that web 2.0 has deep roots in social-network, sociology and medicines as major contribution in research. Web 2.0 has proved best for the researcher as it gives more access to research data and documents [15]. Comparing web 1.0 with web 2.0, the amount of user-generated content was very small with the number of users during web 1.0 being 45 million global users and 250000 sites (mostly read only). It rose to 1 billion+ global users and approximately 80,000,000 sites in web 2.0. The research shows that published

content was higher as compared to user generated content in web 1.0 while on other side, web 2.0 read-write web has a high number of published content, as well as a high number of user generated content [12].

It often happens with technologies that if they start meeting the expectations of users, the technologies begin having trouble from their surroundings. Which in turn causes the slowing down in the growth of the technology or a reduction in overall performance. Some consequences with web 2.0 are as follows:

- The competency of "read-write" in web 2.0 should increase interaction between users on the web.[16]
- Sharing of information between different platforms around the world or communities are still restricted.[12], [12], [12], [16]
- Limited ethical issues regarding using and building of web. There should be proper channel to monitor usage of web.

2.2.3 Web 3.0

Many researches state that the initial footprints of web 3.0 are already here. It took around ten years to make the evolution from web 1.0 to web 2.0, and it might take the same or more time for major changes to take place in restructuring the web. If the next major changes take place in same period (roughly 10 years), in 2015 we have already entered the web 3.0 era. Certainly, people are naming it the internet of things (IoT) and smart home appliances with wireless networks [17].

Web 3.0 is the third generation suggested by John Markoff in 2006. It is the evolutionary and the current web. The main concept of web 3.0 is to arrange data and connect them in order to achieve findings that are more essential and is reusable in different applications. It is also known as executable and semantic web. The inventor of www thought up the semantic web. There is a devoted team working at the World Wide Web consortium (W3C) for the continuous integration and improvisation to standardize systems that are already being used or developed.[12]

Education and research

Web 3.0 is playing a vital role in education and the research field. Web 3.0 is known as a technically advance medium, allowing users to read, write, execute and also

allowing machines to be able to produce outputs which are till the current time only expected of humans. To facilitate learning and education, Web 2.0 and Web 3.0 have created various web based tools and technologies [18]. Universities and educational institute around the world offer electronic services, such as for online education, admissions, interacting with other institute candidates or students. In order to facilitate lifelong learning teachers provides notes, handouts, and other content online and create web-based systems.[18]

Now that we have some idea about web 3.0, we can assume how web 3.0 would be in future or how it will be in coming days. It would be a guessing game predicting web 3.0, a drastic change in web would be how we use it now, and the evolution will take place accordingly in general or in technology. In order to understand more we have a look on some characteristics and scenarios.[17]

Intelligence:
Intelligence is one of the essential feature of web 3.0. Many individuals consider advance artificial intelligence using would be great innovation on web. Web 3.0 era is the period of human interaction intelligent discussion [10]. Direct input of natural language in web 3.0 is also possible. It will help in instant analysis and ideal results achieved, which will save a lot of time. It will save a lot of time in a way, we will receive idyllic outcome, and user does not need to search, match and choose result in enormous amount of data.[18]
In order, applications and websites to work intelligently different Artificial Intelligence (AI) tools integrated like machine learning-Bonsai, kaggle, paxdata, neural networks, Yhat, fuzzy-AI etc. An AI web might be a virtual assistant and many of them support natural language. It means users can communicate in their own languages with other users, it means we can ask some assistance in native language and it will take some mandatory part, convert it and do the task for us like, sending email, make call, make an appointment etc.[17]

Virtualization:
Virtualization is self-explanatory word and its one of the important aspect or characteristic of web 3.0. Web 3.0 would be high-end 3D graphics and high-speed internet bandwidth.[18] These sources would use to create virtualization. The use of

3D designs for web 3.0 in different service, application and websites. The future of virtualization in web 3.0 has an example of "second life".[10]

Personalization:

Personalization is another characteristic of web 3.0. It takes user preferences, as major concern in design that is also the reason user experience considered very important now days. It means everyone will get own preferences whether its individual or personal, it would be consider in various activities information handling, search content presented according to the interest of user. Core technology for web 3.0 personalization would be semantic web.[10], [18]

Semantic Web:

Semantic web is easier and proficient way provided by W3C to find, share and merge information and data from different sources. [19]

 "The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries"

Semantic web is mainly about two things. One is common formats for incorporating and grouping of data from various sources. Where on the real Web primarily focused on the exchange of documents. The language for recording also matters it will reflect how data is relate to real world entities. It allows a machine or person to connect at initial starting point at one database and it continues the connection in endless number of databases, wires do not connect them but being same thing.[12], [19]

Semantic web initially intended the system that allow machines to understand complex requests by humans according to their meaning [13]. Tim Berners-Lee stated for semantic web as follows:

"If HTML and the Web made all the online documents look like one huge book, RDF, schema, and inference languages will make all the data in the world look like one huge database" [20].

A layered architecture proposed by Tim Berners-Lee for semantic web with various variation mainly represented by diagram.
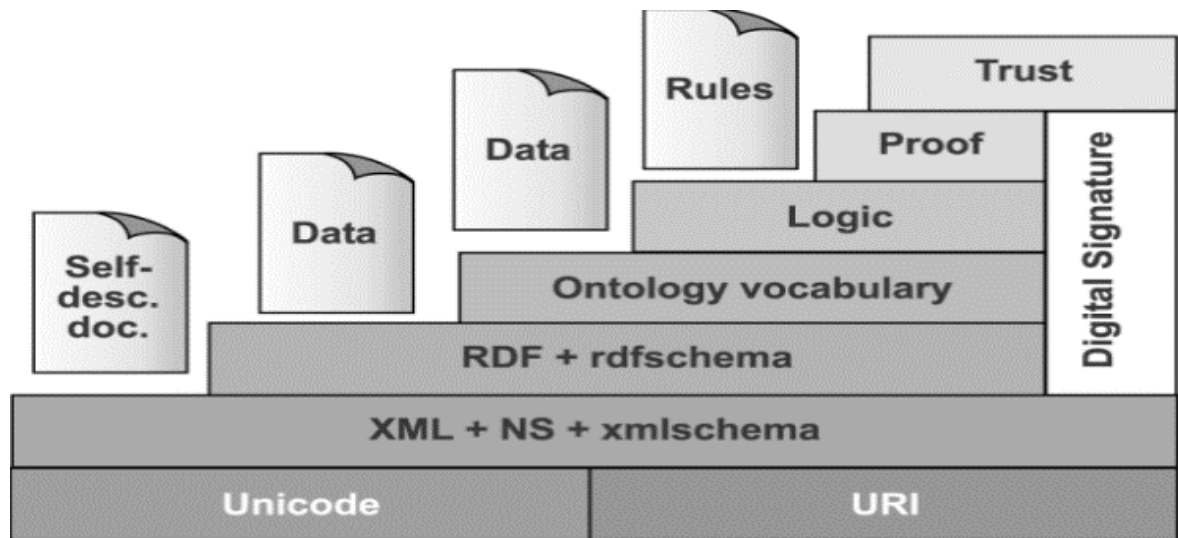
Figure 5. Semantic Web Layered Architecture (Jouis, Biskri et al. 2012a)

Semantic web's development take place in steps, each step builds a layer on the top of other.Figure1 shows the levels of semantic web. It describes the vision and design of semantic web.[12]

URI and Unicode: Unicode is international standard for encoding in order to use with diverse scripts and languages it assigns unique identity or numeric value to each letter, symbol or digit that applies on all diverse programs and platforms. URI is an identifier for a particular resource. The purpose of URI and Unicode is facilitate the mechanism of unique identification within the stack of language for semantic web [21].

Extensible Markup Language (XML): In the world of computing Extensible Markup Language (XML) is a language that defining rules for encrypting documents in such a format that are human readable and machine-readable. It is a markup language and developed by w3c team.XML is open source and easily available for everyone, it does not have any pre-defined mechanism to communicate user new tags meaning [12], [12], [12], [21].

Resource-Description-Framework (RDF): RDF is the meta-data data model. It is part of w3c family. It is standard model for data exchange over the web. One characteristic of RDF is data merging even if the basic schema is different, specifically it support progression in schema with time without the need of all the data users to be changed. [22].

RDF-Schema: RDF schema provides a standard and basic method for RDF models. It is accompanying by several supporting document, which explains the abstract syntax and concept of RDF. This document intended to provide a clear requirement of RDF Schema to those who find the formal semantics specification intimidating. [12], [23].

Ontology: The word ontology is use to refer for complex or formal collection of terms. The Semantic Web relies deeply on formal ontologies to structure data for wide-ranging and transportable machine understanding. Thus, the production of ontologies factors mainly in the Semantic Web's success [24]. Ontology also stated as collection of terms used to explain a particular domain with the capability to intervention [12].

Logic-Layer:

Logic layer is essential part of the web, logic layer means applying rules and policies. Addition of logic layer into web means making inferences by using rules. Rules are a means of stating policies, commercial processes, contracts etc. Many researches focus on using Monotonic logic in semantic web for layered development. It does not provide any mechanism for demonstrating or handling incomplete or conflicting information respectively [25].

Proof-Layer:

It implicates the real logic process also depiction of proof in web language as well as proof validation [12]. Provenance information will be progressively significant at this abstraction level and perhaps the key to creating trust mechanisms and innovative retrieval techniques in decentralized applications [26].

Trust-Layer:

Trust layer is the top layer in layer cake shows in fig 1. It evaluates with proof layer to whether to trust the applications given proof or not [12]. It will occur through digital signature and other kind of information depending on recommendation by trusted sources or on rating and other authorization agencies and consumer bodies. [1]

Semantic web is mainly for developing relations to connect interrelated data. It is not just restricted to publish content or data on the web [12]. In order to connect and publish data on web in 2007 couple of rules were introduce by Berners-Lee, as linked data principle [1].

- Use URIs as label for things
- Usage of HTTP URIs to look up those labels
- Provide useful information, using the standards (RDF) by look up a URI
- To discover more things, include links to other URIs

According to Linked data principle, data providers can post their data to a single universal data space by broadcasting data on the web.

According to linked data principle, data providers can add their data to a single universal data space through broadcasting data on the web.

The main Characteristics of Web 3.0 by Nova Spivack are as follows [27].

- Software as a service (SaaS) Business model
- Open source software platform
- Distributed database or also called as worldwide database.
- Personalization web
- Resource pooling
- Intelligent and smart web

Semantic web has many beneficial characteristics, it also face many issues:

- The World Wide Web (W3) has several billions of pages. The data on the web is too vast, duplication in data can occur, and it has not yet been capable to eliminate all semantically replicated terms.
- The user of information intentionally misled by the content producer [12]. The producer want to deceive the user, in order to achieve some personal or private information.
- Personal and confidential data is damage due to various malicious attacks [28].
- In order to monitor and make sure operation efficiency of complicated application and systems, shortage of inexperienced technician can cause loss.

It has been discussed in many researches about web 1.0, 2.0, 3.0 has differences and evolved with the time. All of them has some major differences such as web 1.0 is web

targets on producer web content creativity and its read only web. Web 2.0 focuses on users and producers' content imagination and its read write web. Web 3.0 is executable web and it aims on interconnected data sets. The difference between web 1.0, 2.0, and 3.0 are shown in following table.

	Web 1.0	Web 2.0	Web 3.0
Implementation Year	1989-2005	2005-2016	2016- Present
	Hypertext web	Social web	Semantic web
Introduced by	Tim Berners Lee	Tim O'Reilly, Dale Dougherty	Tim Berners Lee
Type of web	Read only	Read/Write web	Executable web
Number of users	Millions user	Billion users	Trillions+ users
	Echo system	Communication and participation	Self-Understanding
Environment	One-directional	Bi Directional	Multi user virtual environment
Content Management	Organization publish content	Users publish content	Application made by people to interact with publish content
Content Type	Static content	Dynamic content	AI and 3D, the learning web. Web 3.0 is curiously undefined
	Personal Websites	Blog and social profile	Semi Blog, Haystack.
Communication	Message Board	Community portals	Semantic forums
Contacts information	Buddy List, Address book	Online social networks	Semantic social information

Table 1. Comparision Web1.0 , Web 2.0 and Web 3.0

The above table shows the some major difference in all web versions. Currently web 3.0 is the version of latest era and semantic web is used everywhere. Database and matchmaking are major component in today's web applications. As connectivity of data and interaction with it is also major feature of semantic web.

2.3 Matchmaking:

In the process of semantic web services, finding a suitable service offer in order to accomplish the given or required service request, this process is known as the process of matchmaking. For example, a user is looking for the service, which provides financial information, more exactly stock information. In the advanced market of services, various registered services are eagerly offering to supply information about stock quotes. Different services can be the stock market names, stock prices, stock symbol, stock ticker information etc. [3].
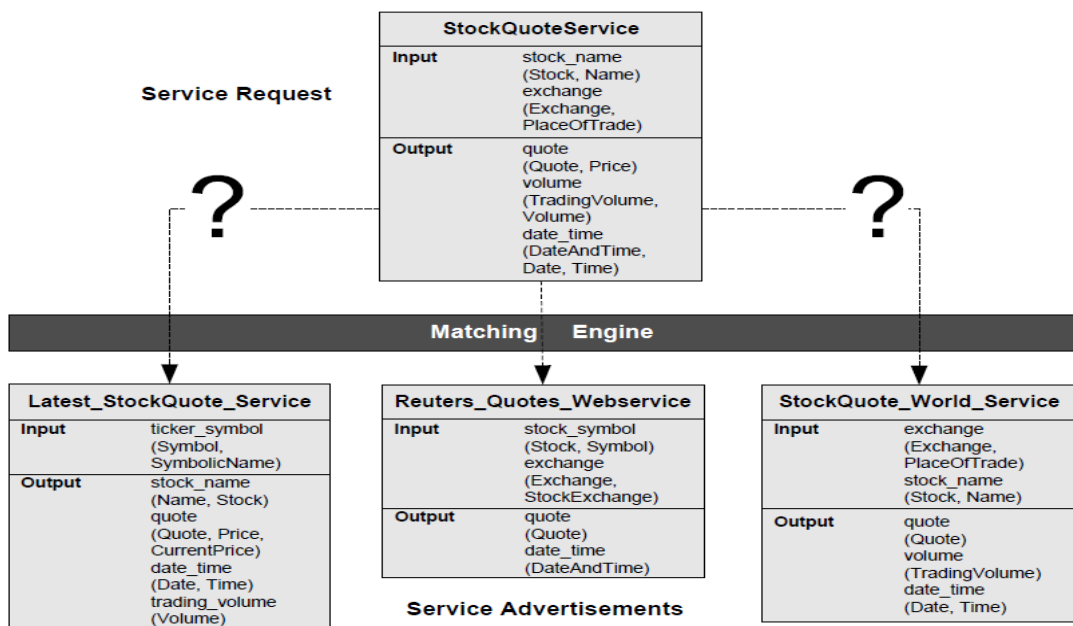
Figure 6. Example of Web Services Match Making (Schulte 2010)

The Figure 6 shows the example of finding ideal service that meet the requirement of the request and provides stock quotes. The service request portrayed on the top of image in top section, the request is explained according to syntax description and names as well as semantic concepts, showing the input and output of the desired service. In the bottom section of the image showed service offered.

The focus of this thesis will be on the functional dimension and semantic matchmaking between services. In this perspective of thesis, three discrete steps can divide matchmaking as follows.

Categorize the data items to be matched:

It is important to define beforehand which information and component of web service to be matched with the service request , in order to fulfil the required request.

Measure Similarities:

Once the data entities that are going to be matched have been recognized, it is mandatory to determine their resemblances.

Matching:

First step is assigning similarities in compound components, second is finding actually the best match component.

Web Services are rapidly growing over the internet, which offer traditional and easy communication for software that are discovered dynamically. Matchmaking of Web Services becomes extremely significant and stimulating [12].

2.4 Related Work

In order to demonstrate the growth beyond the state of the art and including newest feature and ideas. Many research and development teams around the world have worked on different matchmaking tools. The tools and technologies have been generated using various third party services. Such relevant services that are resemble to the current application developed for CST matchmaking are discussed some of them are as follows [29].

2.4.1 Arctic Startup:

The arctic startup is recent growing startup that believe in the working together. They believe great societies are built by collaborative work it distinguish us from Animal kingdom. The aim they want to achieve is unity via associations. Encouraging entrepreneurs that continuously working for making world better place for future generation and change the world with their innovation. On other hand, supporting organization and businesses who believe working on innovation, beyond the boundaries and work with entrepreneurs what they believe was possible.

Developing opportunity for business and entrepreneurs Arctic is working as a bridge and providing entrepreneurs a platform or workspace to work on their ideas, later match them with different businesses. They offer a deal room where they provide

chance both parties to discuss about the opportunities, idea, investment, innovations and working together possibilities. Many entrepreneurs got their matching partner at deal room [30].

2.4.1.1 Deal Room

Arctic15 is famous for proficient networking and deal room is a known place, where magic happens. It a digital platform where participant have the chance to build their profile beforehand and define the services you can offer or what exactly are you looking for. The other best possibility is to reserve 20 minutes meeting most suitable match for you. At the event participant, get the chance to meet and discuss one to one on separate reserved table. Startup have to gear themselves for scaling partners or investments, as well as to seek established companies in order to convince them for investment in their companies. One of the best place to find opportunities, investments and business networks in comfort and privacy. The lawyer are also available there for consultation [31].
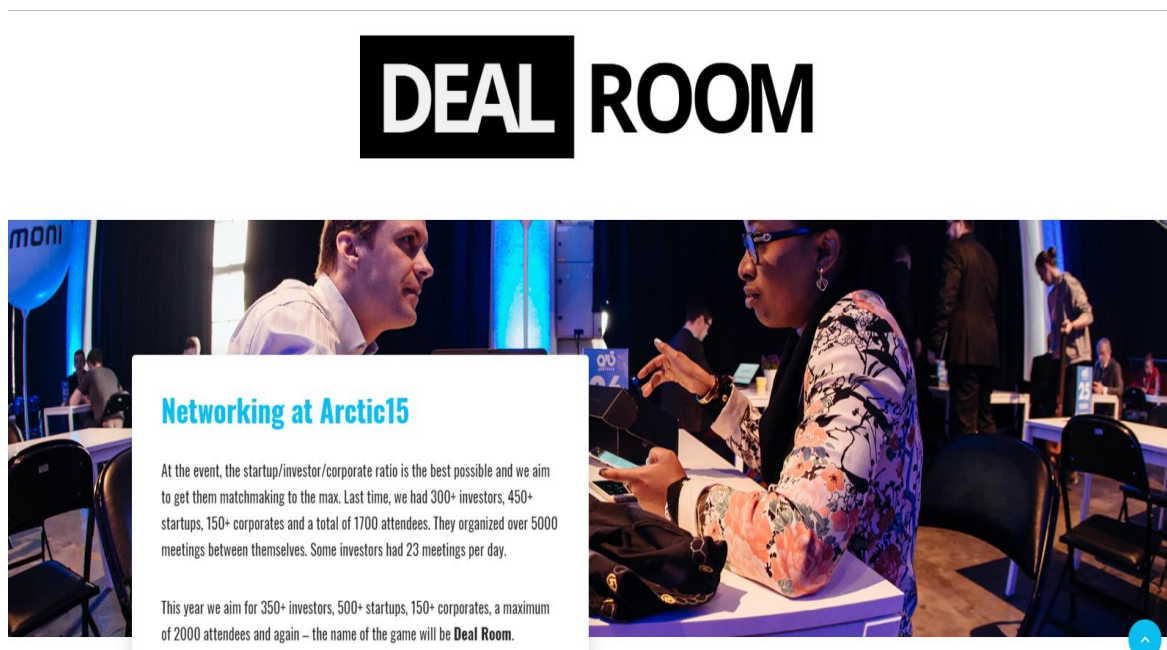
Figure 7. Meeting with suitable match mat Deal Room [31]

2.4.2 Vertical

Vertical is a company that works with different aspects of businesses and individuals.

It is not a startup accelerator nor is a consultancy, the main motive of vertical to improvement by addition. They work and match their clients using five different stages that are as follows.

- Approach
- Campaign
- Shortlist
- Innovate
- Progress

Approach: Initially they know the participant and discuss different aspects, get deep insight of the applicant in order to understand and guide accordingly for business forward. In addition to that, interview with different key stockholders, examining and analyzing the activeness and readiness of organization, to build strategic approach for startup.

Campaign: Once approach is completed, next stage is to promoting via marketing strategies. Vertical activates its network and database in order to set best startup applicant and make sure right match submit their proposal.

Shortlist: Shortlist process is done with intense care, first step is screening and the team look more promising clear the first stage. Second step of screening involves more detail assessment of venture and value propositions. The best team for final decision will get the invitation to selection event.

Innovate: Once everything is set, startup made through final selection and met their match, initially 3-4 months partnership between startup and partners to make sure they meet each other and understand requirements and are able to work together for long term. For that purpose, many mentors are continuously analyzing.

Progress: once all the previous steps are done, vertical is looking forward for providing support in launching the application and deploying prototype, long-term collaborative model between partners and startup [32].

The best thing about vertical they make sure all the matching characteristic in order to work the organization and investor in long run [33].

3 TECHNOLOGIES

3.1 MongoDB

MongoDB is document database with possessing characteristic such as ease of use, flexible and scalability that you need while indexing and querying. MongoDB is an open source and free platform, its first version released on February 11, 2009. It is document oriented and cross platform database program. MongoDB is a database program that is categorized as NOSQL. MongoDB Inc. it is very flexible and stores data in JSON-like documents, which represents structure of data can be changed any time, created in it and fields can fluctuate from documents to documents. It is verified, published and approved under combination of various licenses such as, Apache and server side public license [34].

In application, document model connect or maps to objects for better and ease of working with the code. It has dominant ways to analyze and access data using real time collection indexing and ad-hoc queries. MongoDB is cross platform it makes very flexible for the developer to work on it and can use any language which is supported by MongoDB. There are various famous languages support such as Python, NodeJS, Java, C++, C#, Javascript. Each language has different syntax for connecting to database. Some examples are as follow.

Python Connection string:

```
"# 1. Connect to MongoDB instance running on localhost
    client = pymongo.MongoClient()

# Access the 'separation_target' collection in the 'cst' database
    collection = client.cst.separation_target "
```

Java connection string:

```
"// 2. Connect to MongoDB instance running on localhost
    MongoClient mongoClient = new MongoClient();

// Access database named 'cst' MongoDatabase
    database = mongoClient.getDatabase("cst");
```

```
// Access collection named 'separation_tech' MongoCollection
```
collection = database.getCollection("separation_tech");"

The above strings are the sample for the connection with MongoDB with localhost. Mongoclient is the connection string use to connect with the mongo database. In order to access database, mongoclient.getdatabase is used and name of database inside the quotation and covered with parenthesis. For accessing collection in python, we used "client.cst.separation_target" to access database and collection at the same time.

3.1.1   Architecture of MongoDB:

For most of the organization relational database holds stable place and for good intention. Existing applications are still strongly supported by relational database and meet modern business needs; an widespread ecosystem of tools is supporting them. A big number of qualified labors to maintain and implement these systems are continuously working. Organization are continuously looking for change or alternative to inherit relational infrastructure, due to challenges faced by developers while Building up-to-date application [35].

The nexus architecture
The main design philosophy behind MongoDB is amalgamation of Innovation technologies of NoSQL and critical abilities of relational database. The objective of Mongo Inc is to influence the struggle and work oracle and others are doing over 40 years to make what they are today with relational database. Instead of abandon decades of recognized database maturity, MongoDB continue from where they left off, by merging the work of internet pioneers have been doing to address the challenges of modern application and the important relational database proficiencies [34], [35].

The nexus architecture is shown in figure 7. It shows the combination of relational and NoSQL database. It shows the important aspect of relational database with NoSQL technologies. The main feature of NoSQL is flexibility is which is available in nexus architecture. The MongoDB stores data in Binary encoded-serialization (BSON) format document. BSON is extension of JSON format or can be said it is derived from JSON.
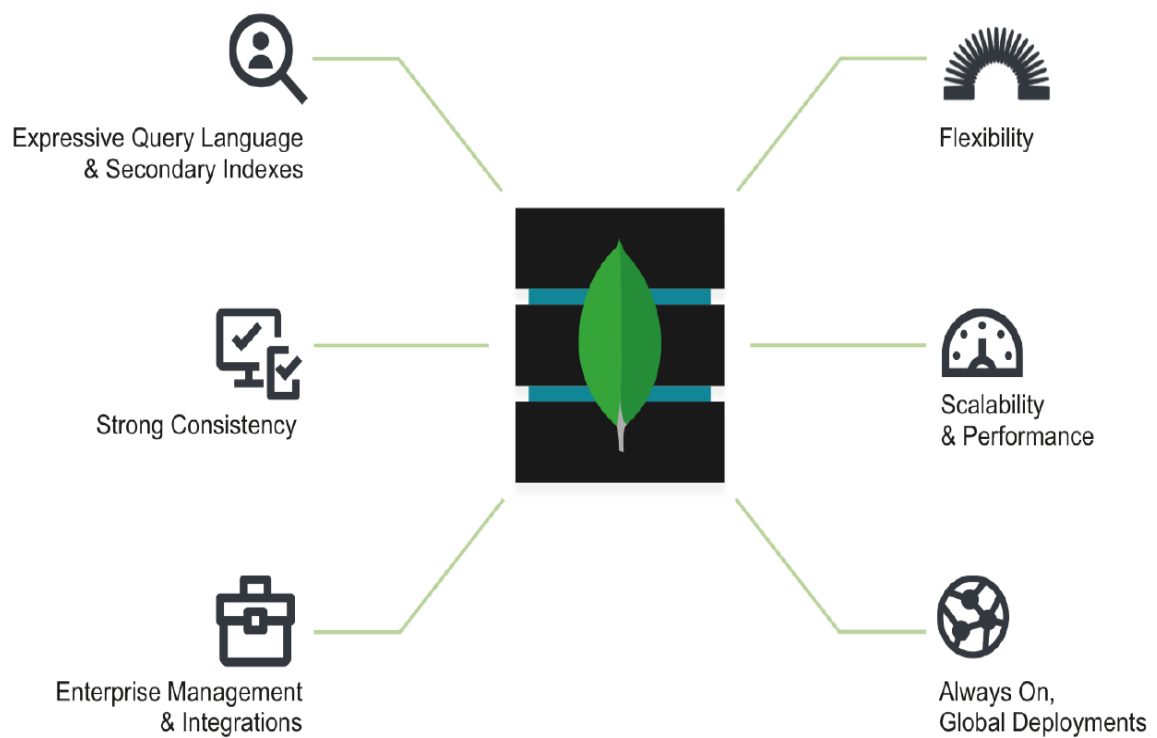
Figure 8.MongoDB Nexus Architecture, blending the best of relational and NoSQL [35]

The main attributes of Nexus architecture that are shown in the figure 7 are as follow:

- **Expressive Query language and secondary indexes**

- **Strong consistency**

- **Enterprise Management and Integrations**

- **Flexible Data Model**

- **Scalability and Performance**

- **Always-On Global Deployments**

3.2 Python

Python is old Programming language, its first version was introduced in 1991 by Guido van Rossum. It is interpreted and top level programming languages, used for general programming purposes. Python is a language, which allows you to work rapidly and build applications and systems efficiently. Python use Object-oriented concepts and dynamic semantics for creating application. It possess top-level data structures joint

dynamic binding and dynamic typing makes it perfect for Rapid application development (RAD) [36].

Code reuse and modularity is encouraged in python, due to support of modules and packages in it. Many developers starting loving python due to it fast productivity, there is rapid cycle of edit, test and debug cycle, no compilation step involved in it.

3.2.1 Python Packaging

The terminology of Packaging is quite simple, Packaging in python is kind of folder which contains files of python. Modules terminology is used for python files. Sometime python developer themselves are unclear about packages, in order to remove this vagueness developer use term "Python package" especially when it is discussed about python packages. Most of the developer dream of writing such a code that it works on most of the platform, but for that, it is always require that a package contain all the dependencies or libraries with it while building a complete package. However, python based release is a made with a objective of it working with python without any difficulty, without boundaries of operating system (OS). Every developer hope that:

- Packager for each targeted operating system which helps it in repackage the work done by developer.
- Dependencies in the package of developer will also repackaged in each system where work is going to be used.
- Clearly defined system dependencies.

Sometime, it is basically impossible, for example, Plone (a CMS completely based on Python) Plone uses many hundreds of small and big libraries, that are not always available for all packing systems or for the system where application of code going to be used. It shows that Plone CMS system must export everything that it require in portable system or application. To overcome this problem it uses "**zc.buildout**" that collects all of its dependencies and generate a portable application, which will run or work on any system within a single directory without adding anything else. This is big success for developer, they just need to use zc.buildout and using python standards describe their own dependencies [37].

How to handle data files, it is another problem. For example, if SQLite database is being used by the application, if the application is kept in package directory, the chances are system will prohibits you to write in that specific part of the tree and application might fail. The above-mentioned problem can be riddled with python library standard package "Distutils". It is standard package, people either use it or live with its flaws, or use more enhanced tools like "Setuptools", it add additional feature above it or on Distutils. Other more advanced installer pip that relies on setuptools.

Usage of Distutils is possible, developer need to add a single file or module to project, named as setup.py. Example of that define standard options as follows:

> **"from distutils.core import setup**
> **setup(name='MyProject', version='1.0', py_modules=['mycode.py'])"**

The module can be run using command

> **"$ python setup.py sdist"**

Same script can be install by using install command.

> **"$ python setup.py install"**

3.2.2 Packaging architecture

Python architecture is distributed in various steps. It is connected via many entities. Everyone interact with setup.py, whether to build. Publish, install or package. The developer of his project via option distributed to function describes the content and for all packaging task uses that file. The same file is used for installation on target system by installer. Further process can be seen in following figure. Which reflects all the relevant connected entities.
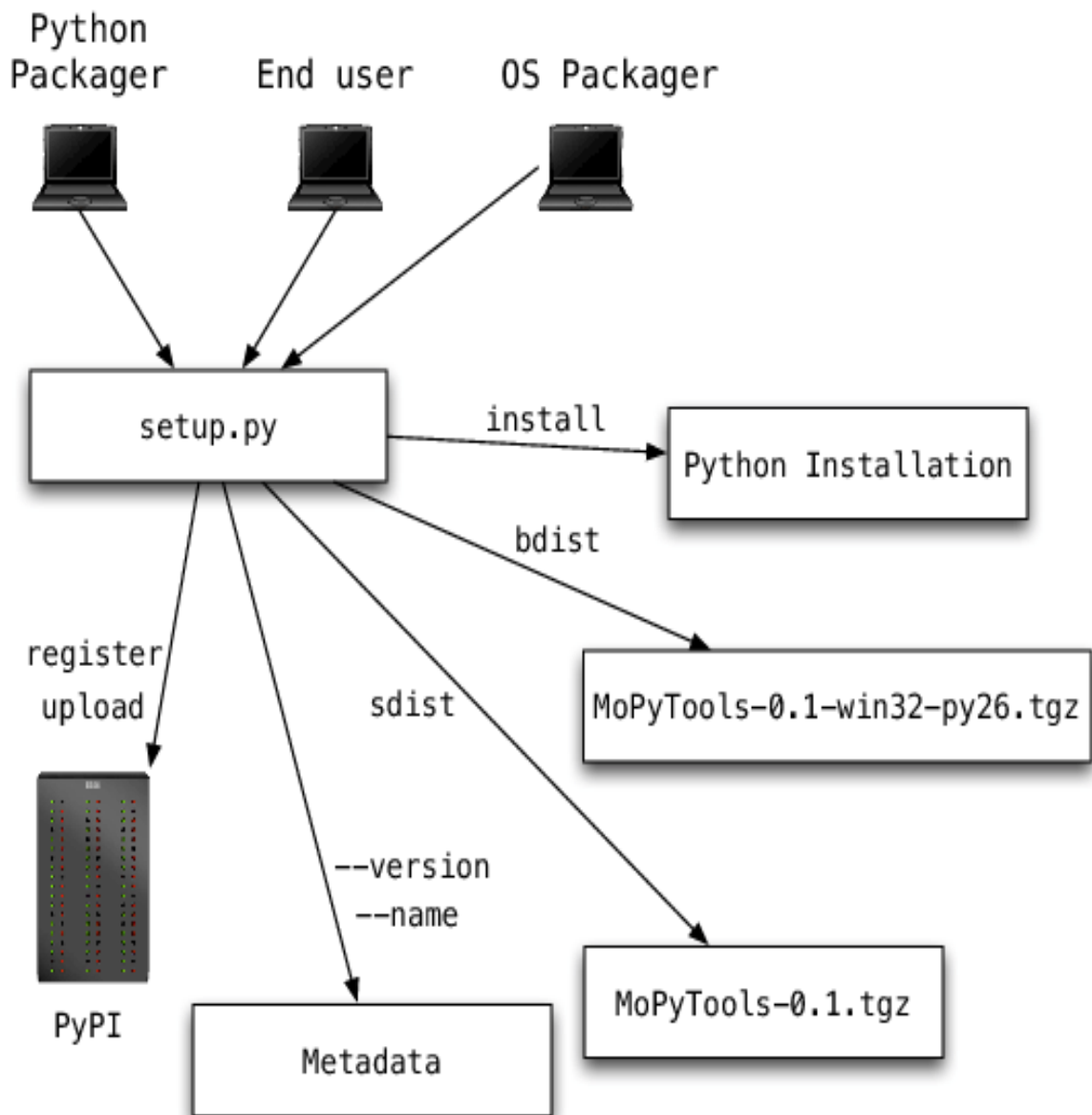
Figure 9. Setup of Packaging

3.2.3 Architecture of PyPI
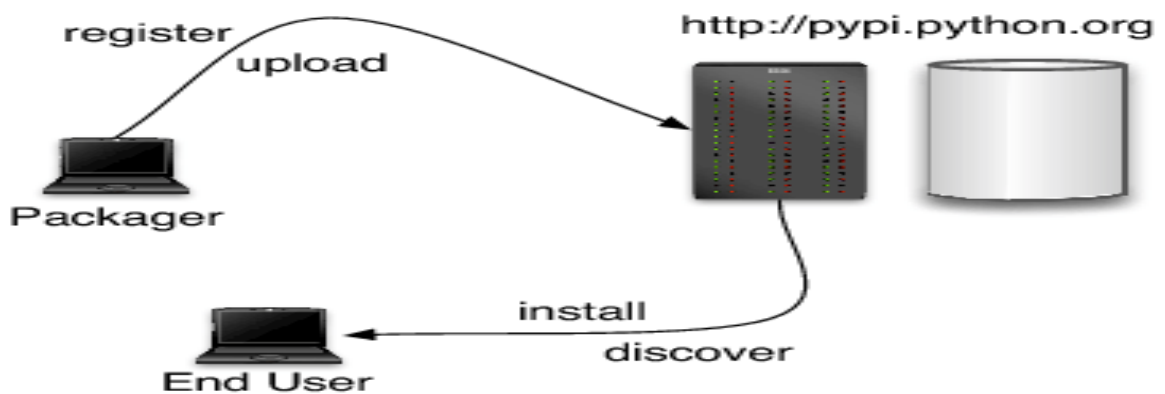
Figure 10. Architecture of PyPI [38]

Central index of python projects is PyPI, people can look up for their projects according to Category or people can register their work [38].

3.2.4 Flask

Flask is a framework that is highly recommended for the developer who wants to create standalone applications. Flask has built-in jinja template, jinja is complete featured template engine. It is BSD licensed and widely used. It has a voluntary incorporated sandbox implementation environment. Example of Jinja template is shown in following code:

```
{% extends "layout.html" %}
{% block body %}
    <ul>
        {% for user in users %}
          <li><a href="{{ user.url }}">{{ user.username }}</a></li>
         {% endfor %}
          </ul>
{% endblock %}
```

Beside of having a built-in template, flask can also choose any other object relational mapping (ORM) or template engine such as Django, CherryPy, Pylon etc [39], [40], [40], [40]. Flask is very beneficial for endpoints, APIs or RESTful services. Developers are free to develop backend the way they like it, but it was mainly intended to design open-ended application and systems [39], [40].

4 IMPLEMENTATION

4.1 Consideration for the mode of research

The mode of research was processed in different parts, which consist of many important factors. First part of implementation was about research and second part consist of development of application and database. Data was one of the main part of this application, arranging the data and managing it accordingly in order to fulfil the requirement of the application. As this was the mandatory requirement of the application as it is a content-based application, which will be presented when user sends any request for the required information. Collecting the data was one of the challenging task during the whole, nowadays open data is very famous for arranging data and available online. The data with open and big data is accessible from everywhere, as it clearly understandable from its name open data means open for all. The data that was required for the idea we were trying to develop was different and the data that was available at open data was not fulfilling our requirement.

4.2 Collection of Data

In order to fulfill the requirement of data for the application, we need to do a lot of research and finding data precisely and specifically. As the information we needed was the different companies main work areas, the technologies they use, what industry they belong to, how does separation work take place for them, separation processes, separation target, what applications they belong, what solution they offer for different chemical removals, etc. It was one part of the data, collecting the information for professors and their expertise was another challenging task. There were some professor who were chosen as initial participant of this application. The data was gathered and arranged in excel sheets.

4.3 Setting up Environment

Data collection is prime part in process of development, after that next step possess setting up the environment. Setting database and managing tables and installation of softwares for the usage of databases and technologies that are part of development such as MongoDB, studio3T for databases and for installation of backend

development, Python environment. In order to work on the code there is usage editor PyCharm and notepad++ plays the role of editor in the process of development. For frontend development, the tools were used CSS, Bootstrap, jQuery, Ajax and HTML. For storage of application and accessing it from anywhere and anytime, Azure cloud storage is used.

4.3.1   Setting up MongoDB

Setting up MongoDB is one of the important task, as it can create problems later if it's not properly installed or setup. While installation there are various option that are needed to choose according to the requirement of the project. One important aspect, which is mandatory for all the operating systems. Such as for windows it requires directory which stores all the data, its known as Data directory. The location of the data directory is in "\Data\db" [41]. All the log files store in this folder and can be easily accessible for administrator. If any error occur, we can check details and find exact error and can be resolved accordingly.

After setting up MongoDB to run it we need to access the MongoDB folder in windows folder program files and run mongod.exe on following link.

"C:\Program Files\MongoDB\Server\3.2\bin\mongod.exe"

This file also can be accessed via command prompt, an example is shown in following figure.8, first we will start command prompt using windows key on keyboard and press Ctrl + Shift + enter. Afterwards go to the MongoDB directory is available and access above mention path, and just write mongod.exe as it is shown in the figure 11. Once mongod start executing, it will read some file and show the execution which is shown in figure12.

Table 2 shows the list of additional and specific component sets to install with MongoDB, in order to install them all together ADDLOCAL argument using one or more of the component shown in Table 2, and can be accessed by commas separation [34].

Figure 11. Example of Starting mongoDB

Figure 12. Execution of Mongod.exe[41]

Component Set	Binaries
Server	Mongod.exe
Router	Mongos.exe
Client	Mongo.exe
Monitoring Tools	Mongostat.exe mongotop.exe
Import Export Tools	mongodump.exe, mongorestore.exe, mongoexport.exe, mongoimport.exe
Miscellaneous Tools	bsondump.exe, mongofiles.exe, mongooplog.exe, mongoperf.exe

Table 2. MongoDB component sets [41]

4.3.2 Setting up Python:

Python has multiple versions current version is Python 3, it is vastly preferred to use python 3 over python 2 version. Upgrading from python 2 to 3 is simple and easy, just need to follow some simple steps for upgradation. Once python is installed on the system, python is easily available online on their official website, which is as follows:

https://www.python.org/download.

Each setup processing is quite different according to the Operating systems, for application creation windows operating system was used.

It is recommended by some user manuals and guides to install tools and libraries before start building applications for real use. Specifically install pip, virtual environment and setup tools before start building application it helps a lot, make things easy and help in using other third party libraries. Figure 10 shows during installation process make sure to check the box on **Add python to PATH,** as interpreter make sure to place in the execution Path.

Python's two most vital third party packages pip and setup tools, they accommodate in installation, uninstallation and download product with a single command. Example of pip installation as follows:

"python -m pip install -U pip"

Installation of "pipenv" is also necessary, as it helps in managing virtual environment and dependencies. Virtual environment is important tool, which helps in keeping dependencies needed by various project, for python virtual environment at different locations [42].

Figure 13 Installation Of Python and Adding to PATH [43]

Figure13shows that it can be changed to specific users that means if it is uncheck on Install launcher for all users, it will install some component for specific users. Environment variable is one of the important factor for accessing. Environment variable are variable which describe in which environment program is running. Using python in windows environment variable possess a name and value [44].
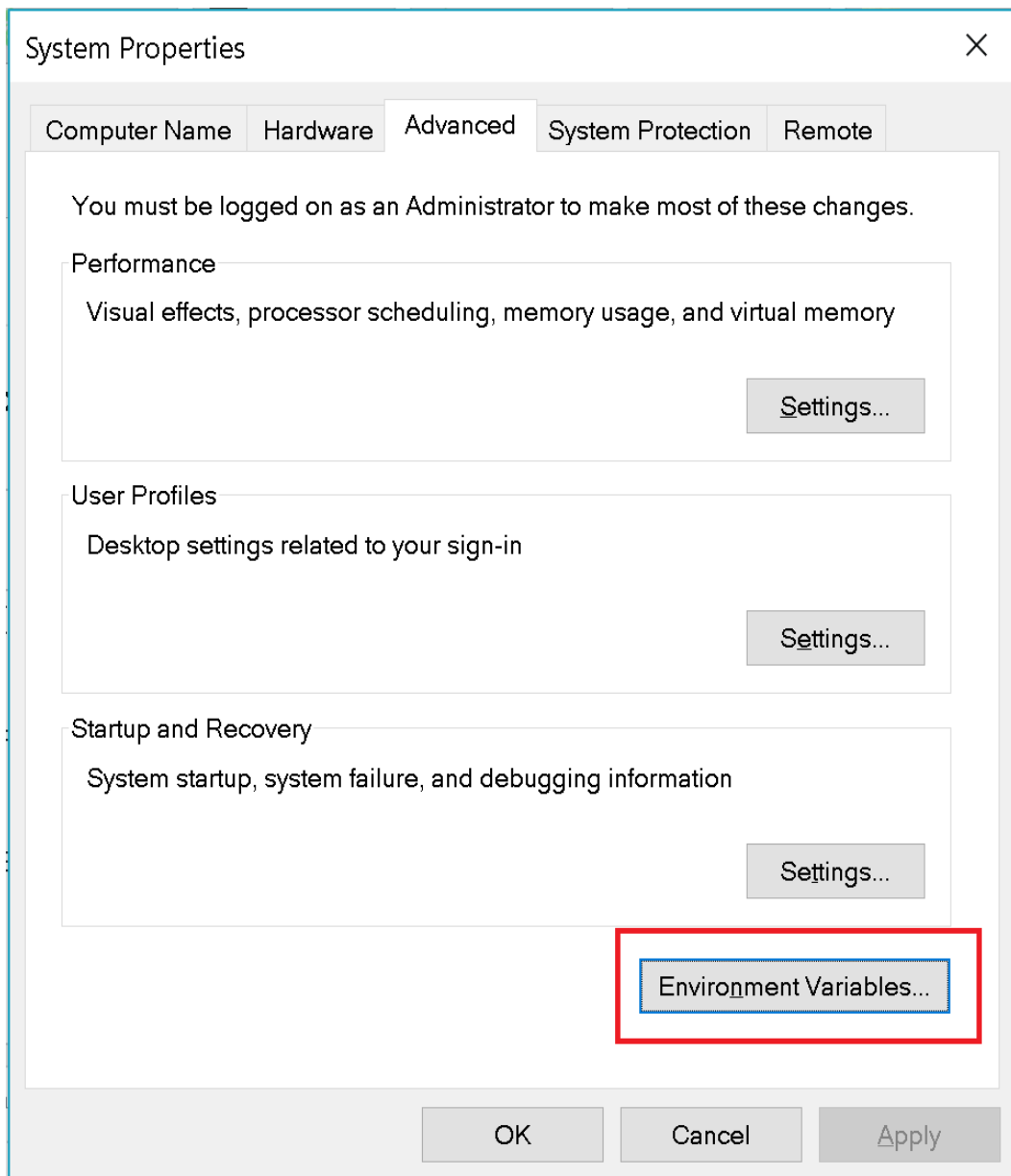
Figure 14. Environment variable windows [44]

In windows directory windows environment variable can be accessed via desktop, right click on computer icon, select properties from context menu, afterwards choose advance system setting and click on environment variable, via edit system variable, specify the value of the environment variable and update PATH value [44].

Figure 14 shows the post window after click on environment variable, as it is shown in figure15; new, edit and delete button. For purpose of adding environment variable, click on new button and add path and name of environment variable. In order to edit previously mentioned Path and value, we need to use edit button [45].
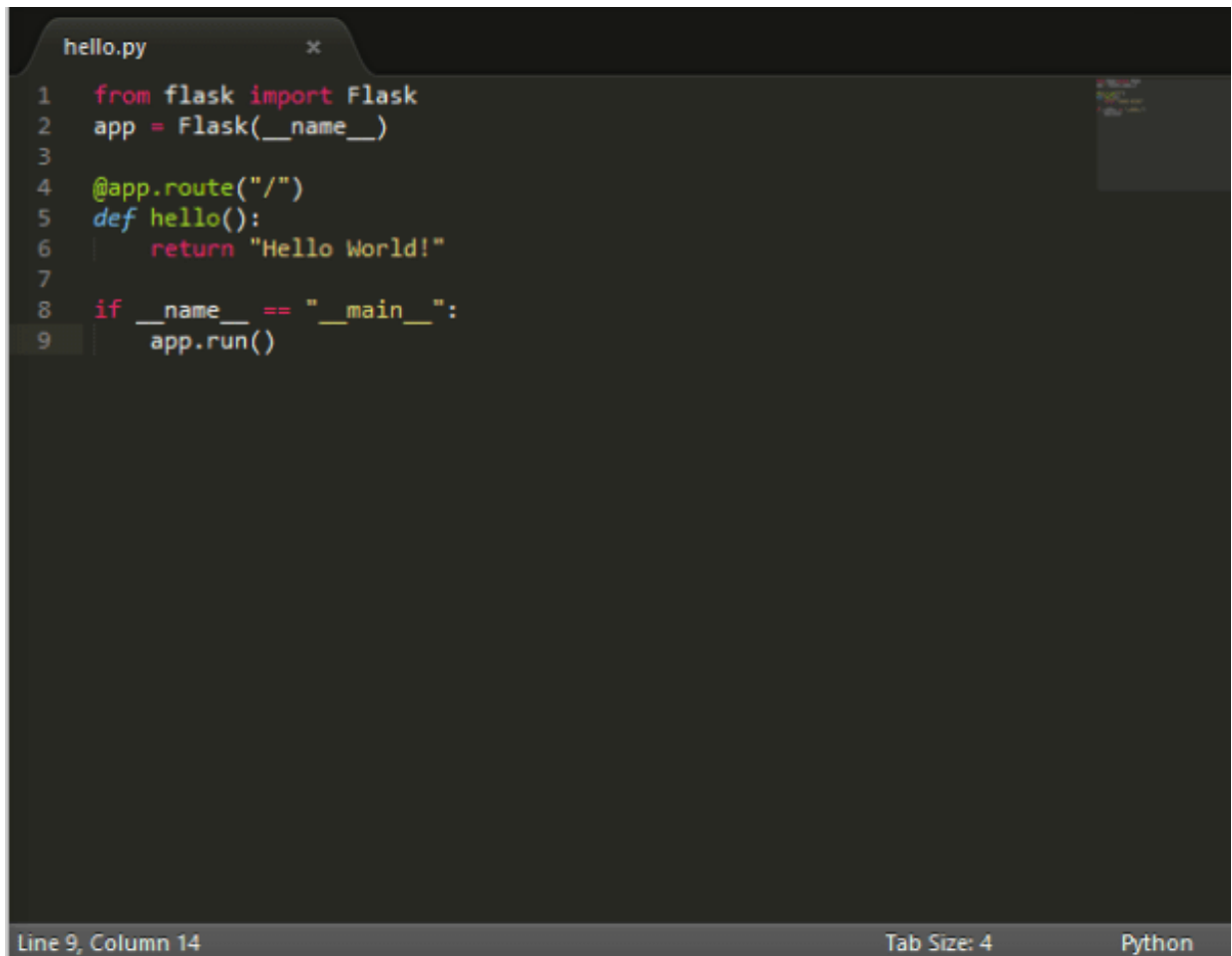
Figure 15. Assigning the value of Environment variable [45]

4.3.3 Setting up Flask

After installation of database and python, it is time to setup flask in virtual environment and application code. After downloading of flask in the virtual environment and installing it using PIP installation. Before installing Flask, make sure availability of virtual environment is active and working. Flask can be installed by writing simple one line code in shell and virtual space on the system the code for installation is in few simple steps, which are shown as follows:

”$ pip install Flask”

Flask is installed and ready to use, now open editor of your choice and start working on application; basic hello world example from flask website is shown as follows.

```python
from flask import Flask
app = Flask(__name__)

@app.route("/")
def hello():
    return "Hello World!"

if __name__ == "__main__":
    app.run()
```

Figure 16. Example of Hello world and Flask use (TImbo 2014).

The hello world code is basic code is mostly being used for all languages around the world, it is used in order to test the initial code and make sure that everything related to it Is working accordingly. Such as Virtual environment, interpreter, modules etc. As flask is imported using the python script "from flask import Flask" app.run is method to run the code, we run the code using virtual environment or editor has the option of running the code in it. A example of running code is shown below:

```
     _speedups.obj : warning LNK4197: export 'init_speedups' specified multiple times; usi
ng first specification
     Creating library build\temp.win-amd64-2.7\Release\markupsafe\_speedups.lib and obj
ect build\temp.win-amd64-2.7\Release\markupsafe\_speedups.exp
  Running setup.py install for itsdangerous
  Running setup.py install for Jinja2

  Running setup.py install for flask
Successfully installed Jinja2-2.7.3 Werkzeug-0.10.4 flask-0.10.1 itsdangerous-0.24 markup
safe-0.23

(HelloWold) C:\Users\tireilly\dev\HelloWold>dir
 Volume in drive C is OSDisk
 Volume Serial Number is 54D1-68DF

 Directory of C:\Users\tireilly\dev\HelloWold

03/31/2015  12:58 PM    <DIR>          .
03/31/2015  12:58 PM    <DIR>          ..
               0 File(s)              0 bytes
               2 Dir(s)  100,612,603,904 bytes free

(HelloWold) C:\Users\tireilly\dev\HelloWold>dir
 Volume in drive C is OSDisk
 Volume Serial Number is 54D1-68DF

 Directory of C:\Users\tireilly\dev\HelloWold

03/31/2015  01:18 PM    <DIR>          .
03/31/2015  01:18 PM    <DIR>          ..
03/31/2015  01:18 PM                151 hello.py
               1 File(s)            151 bytes
               2 Dir(s)  100,612,603,904 bytes free

(HelloWold) C:\Users\tireilly\dev\HelloWold>python hello.py
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
127.0.0.1 - - [31/Mar/2015 13:25:52] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [31/Mar/2015 13:25:52] "GET /favicon.ico HTTP/1.1" 404 -

(HelloWold) C:\Users\tireilly\dev\HelloWold>
```

Writing only this line "python hello.py" and press enter will execute the code in python

Figure 17. Example how to execute the code [47]

and give us and IP address, which is actually a localhost link, when we run that link it will display the execution of code or the results we are expecting from the code.

4.3.4   Design and Development:

The methodology that was followed for the development of project was "Agile". The purpose of considering agile is, its development focuses on rapid life cycle. The main process of agile are described as follow:

- **Concept**
  Projects are planned and prioritized
- **Inception**
  Initial requirement and environment are discussed, team members are chosen for various tasks and funding is set accordingly.

- **Construction or Iteration**

  The production team works on task in order to deliver working products, parallel considering requirements and receiving feedback.

- **Release**

  This phase consist of testing, quality assurance (QA), trainings internal and external, and final release after successful iteration into production.

- **Production**

  Continuous support for the developed software, working and fixing bugs and continuous development of improved versions.

- **Retirement**

  End of product development life activities, it also includes customer migration and report.

The agile software development life cycle view represent as follows:

Figure 18. Agile process Flow [46]

The agile life cycle follows iteration process and each iteration has different feedback and bugs to solve and improve. Documentation is working software's supporting element, which is available to the customer until final software product is ready. Customer can update the document until it does not meet requirement of customers. During whole software development life cycle, multiple iteration takes place, each iteration has different workflow according to the iteration reviews, and changes need to improvise. A standard iteration process workflow can idealized as follow:
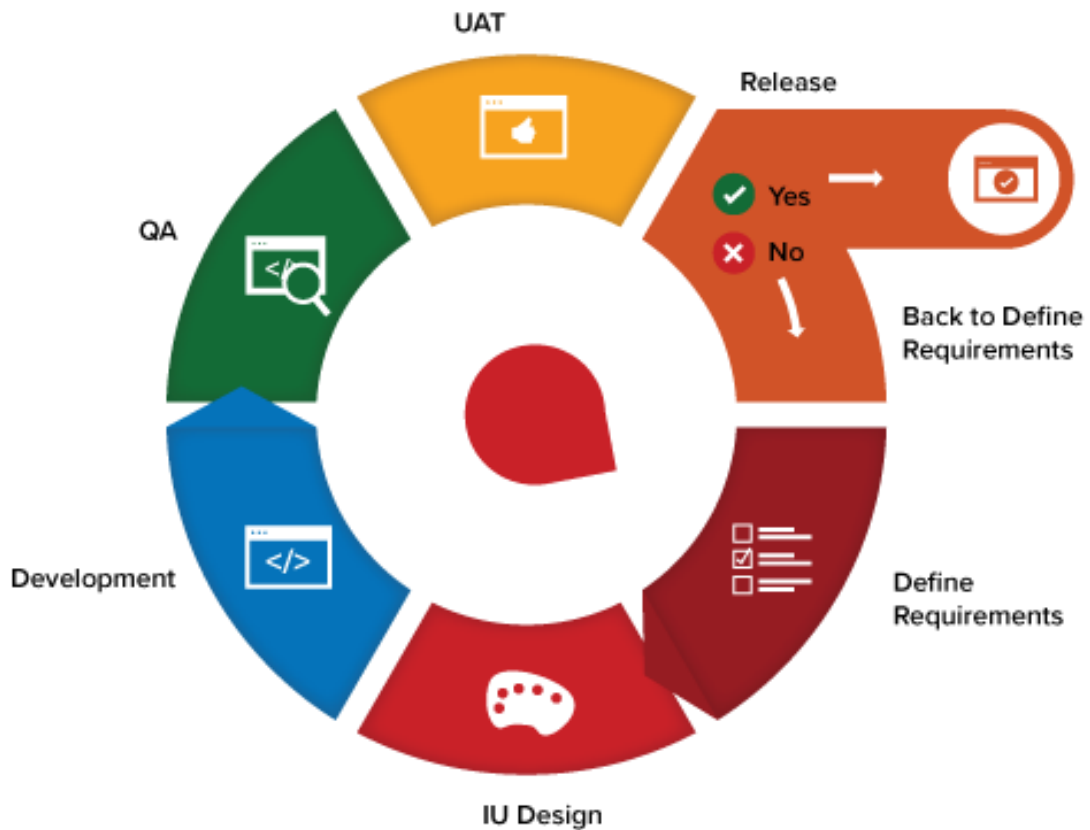
Figure 19. Iteration Process Flow [46]

- **Requirement**

  Iteration has its own requirements and it has to be defined before iteration start depending on the sprint backlog, backlog and feedback from customers.

- **Development**

  Based on the pervious phase where requirement based on iteration has been defined, design and development take place.

- **Testing**

  Testing phase require more activeness and care while testing software according to documentation, internal and external trainings in order to quality assurance.

- **Delivery**

  Once product development completes the iteration and it is ready to deploy or deliver to the customer.

- **Feedback**

The user will start testing it and share their experience and feedback about the product, the feedback is used in backlog, sprint backlog so the next iteration take place according and requirements of iteration will be made out of it.

The above mention workflow followed during the development process of CST matchmaking database. There were requirement and continuous testing of design and development from the project manager Samantha Kiljunen, afterward feedback and changes takes place accordingly. The design and development for the CST matchmaking database are as follows.
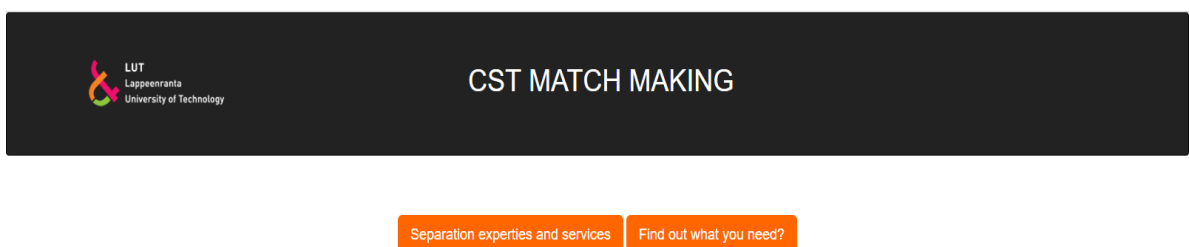

Figure 20. Main landing page of CST matchmaking

There are TwoButtons on the main screen, first labelled "**Separation expertise and services" the** other one is "**find out what you need?"** both buttons work differently and reflect different results. The main reason for creating and accessing further data using only two buttons is ease of use for customers or users. The simple two button concept help customers to find out what they are looking for and where they can find the solution totheir problems. "**Find out what you need"** button helps customers if they do not know the type of separation they are looking for only the phenomen they want to solve.. User will simply follow the user interface, once they choose any of the buttons or click they will be processed with further options. Which are shown in next figure.

Figure 21. Options after selecting separation expreties and services

As it is shown in Figure 21 when separation expertise and services is selected two further options appear. Expert and separation technology provide the user with a drop-down menue to select differernt options.

Professor	Industry	Application	Seperation target	Seperation Technology	University	Position	Country	
Antti Häkkinen	Side streams	Green Liquor dregs	Hazardous metals	Filtration	Lappeenranta University Of Technology	Professor	Finland	Details
Antti Häkkinen	Side streams	Green Liquor dregs	Hazardous metals	Hydrocyclone	Lappeenranta University Of Technology	Professor	Finland	Details
Antti Häkkinen	Side streams	Green Liquor dregs	Hazardous metals	Chelating agents	Lappeenranta University Of Technology	Professor	Finland	Details
Antti Häkkinen	Metals & mining	Bauxite		Filtration	Lappeenranta University Of Technology	Professor	Finland	Details
Antti Häkkinen	Pulp mills	Boiler Flyash	Trace metals	Crystallisation	Lappeenranta University Of Technology	Professor	Finland	Details
Antti Häkkinen	Pulp mills	Boiler Flyash	Trace metals	Solid-liquid	Lappeenranta University Of Technology	Professor	Finland	Details
Antti Häkkinen	Pulp mills	Green Liquor dregs	Trace metals	Filtration	Lappeenranta University Of Technology	Professor	Finland	Details
Antti Häkkinen	Metals & mining	Bauxite	Solute recovery	Filtration	Lappeenranta University Of	Professor	Finland	Details

Figure 22.Experts result

Selecting experts as shown in above figure will give users the direct access to the list of all experts. The user can check the expert's name, the industries they work in, applications they are expert in, separation target and separation technology that can be used for separation, country and profession are also shown, and so it can be checked with country or profession of the expert. If any user want to contact the expert, they will get contact information via the detail button.
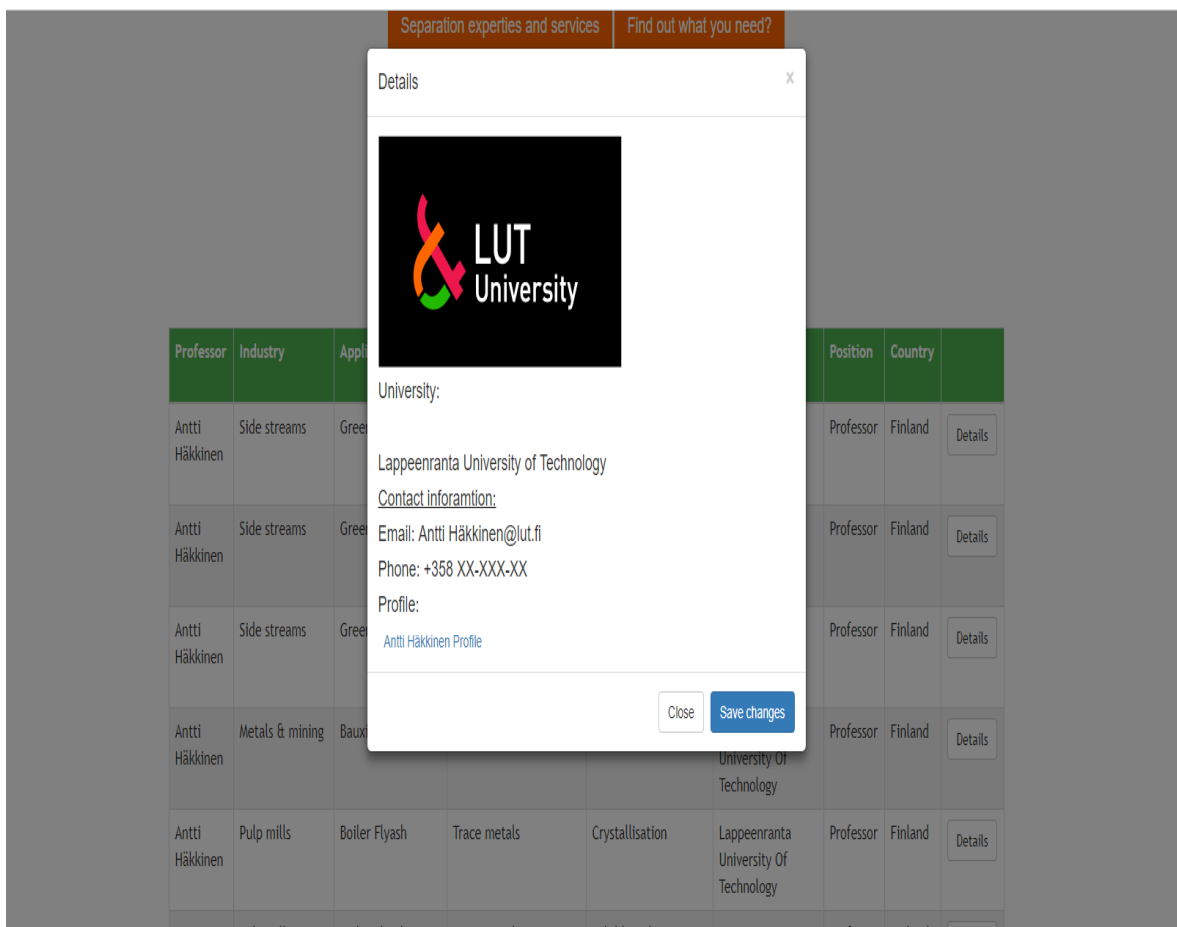

Figure 23. Detail page of experts

The information ishown in the above picture,shows the university name, email address of the expert, contact number, and there is also an option of viewing the profile of that specific expert. Contact information of the users have been provided with the confirmation and acceptance of the experts, due to new data confidentiality law phone number haven't been kept visible.

Figure 24. Separation Technology
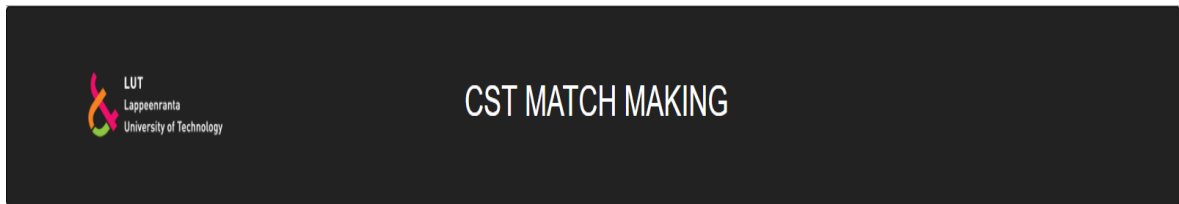
The next button in the separation experts and services is separation technology. Selecting this buttonresults in a pop-up of three more options for the users, which help them in selecting from the following options:

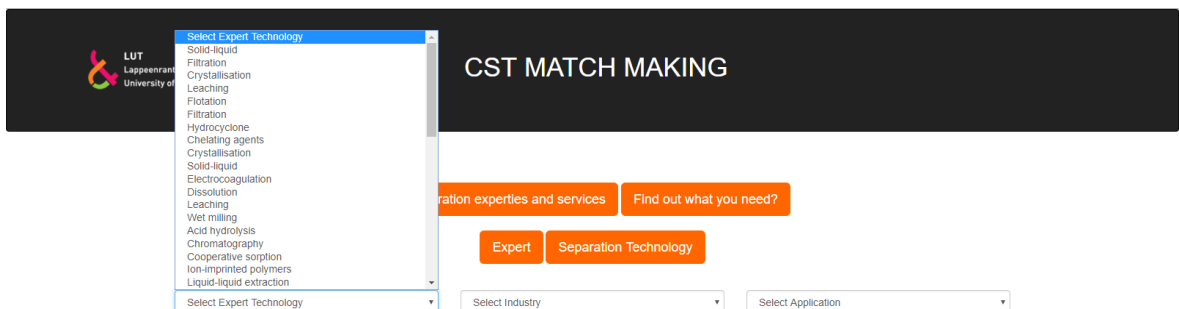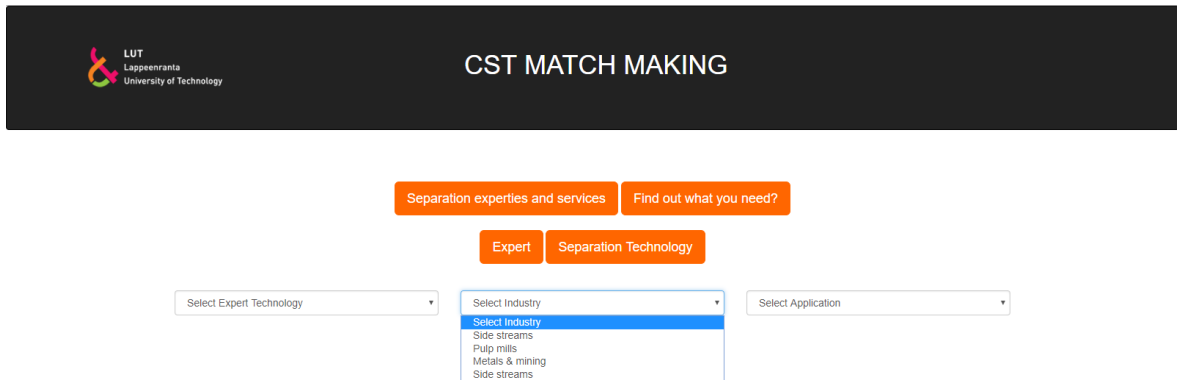- **Expert technology**

- **Industry**

- **Application**

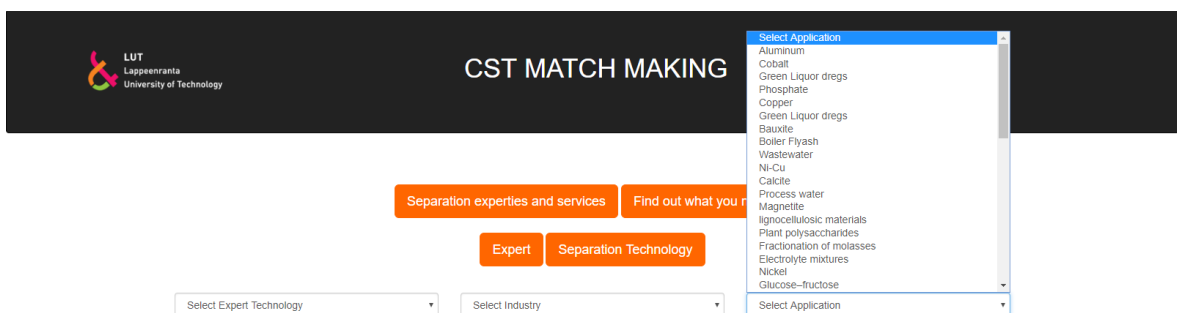**Figure 25. Expert Technologies**

As it can be seen in Figure 25 various technologies are shown on selecting the dropdown option of 'the select expert technology button'. We need to selectone option from this category

Figure 26. Select industry

The second drop down gives the option for selecting type of industry, the sample data of industry is shown in Figure 26

Figure 27. Select Application

The third and last option reflect the data of application in above figure, once we select all the drop down options it will reflect in the search button on the page. After clicking the search button, we will get the results. Which is shown in the Figure 28 showing all the matching results, that meet the search criteria requested by the user by choosing drop down menus. In all the search results, we can also see search filter options that show just above the result table. It helps the user to scrutinize or increase filtering out of that search results and search within specific results. For example, if some users want to check only with expert name they just need to write down the professor name in the filer bar and it will show the results accordingly in the same table.

Professor	Industry	Application	Seperation target	Seperation Technology	University	Position	Country
Antti Häkkinen	Metals & mining	Wastewater	Sulphate removal	Electrocoagulation	lappeenranta university of technology	Professor	Finland
Antti Häkkinen	Metals & mining	Wastewater	Sulfate removal	Electrocoagulation	lappeenranta university of technology	Professor	Finland
Antti Häkkinen	Metals & mining	Wastewater	Sulphate removal	Electrocoagulation	lappeenranta university of technology	Professor	Finland
Antti Häkkinen	Metals & mining	Wastewater	Sulphate removal	Electrocoagulation	lappeenranta university of technology	Professor	Finland

Figure 28. Results on choosing all three options

Separation expertise and services module is completed here, now the next module that is 'Find out what you need?' it will direct the user to select what separation target and chemical they want to remove from their products or process
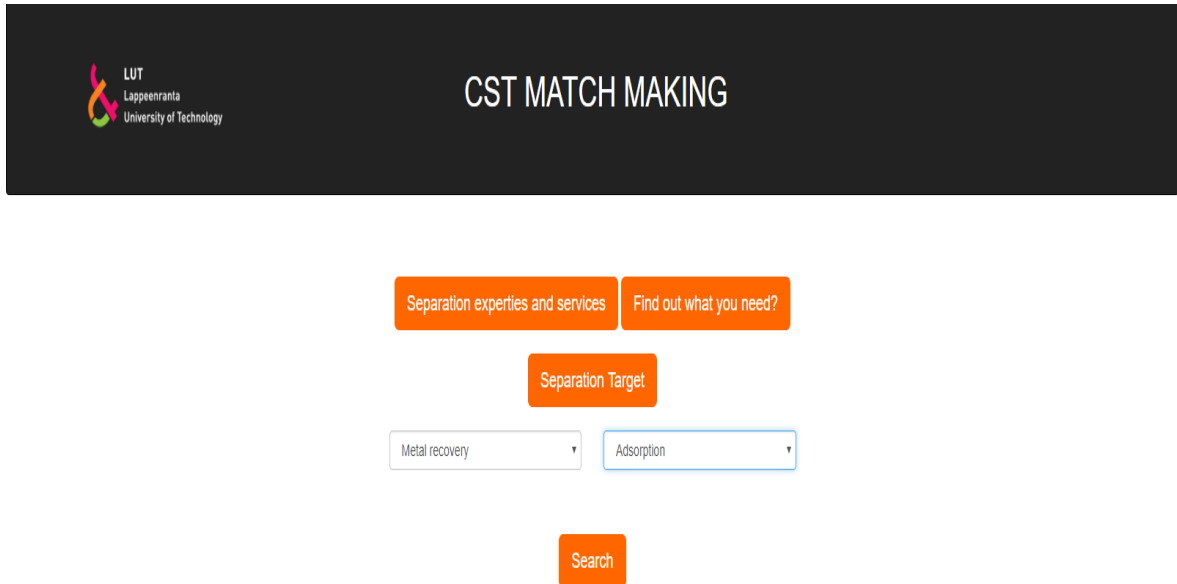
Figure 29. Find out what you need Button and options

The above figure shows the option obtained after selecting find what you need button. Separation target is the button that helps users in obtaining alternative separation options for a challenge they want to solve.

Figure 30. Separation target option

The separation target button further distributes in two options, one is select target separation and other option is select target technology. Once the choice is made from both the drop down a search button will appear. The result will be shown after hitting click on the search button .

Figure 31. Result of Separation target.

5 TESTING AND EVALUATION

In the Chapter 4, the implementation is presented and the complete process that takes place during the development of the application. This is the initial version of testing for further modules. A survey was conducted during the presentation of the application, which was conducted at the CST event in Helsinki August 2018. The application was tested by various users, in order to get feedback and future development work needed a survey was provided to all users to fill in on completion of testing of the database. . The survey questions and their feedback are as follows:

 Question 1: Which Industries are you working or have worked?
Feedback:
- Metals and Mining
- Mining, Pulp & Paper
- Marketing
- Marketing
- Additive and laser based manufacturing
- Forest, Chemicals
- System Analysis
- Chemical Engineering
- Measurement technology
- ICT and Water
- ICT and IT Services
- Mining , Bio and Water
- Higher education communication, Forest industry, Packaging board
- Chemical and process industries, Pharmaceutical, Metal and mining
- IT

Question 2: What is your current Designation?
Feedback:
- Professor
- R&D Manager
- Web Developer

- Designer
- Additive Manufacturing
- Principal Science
- Research assistant
- Student
- CEO Labkotec Oy
- Project Manager
- HR and Marketing
- Professor
- Communication specialist
- Professor
- CTO

Question 3: How would you rate this application? From the ease and usage point of view?

Feedback:

Question 4: Would you choose this application over Google search?

Feedback:
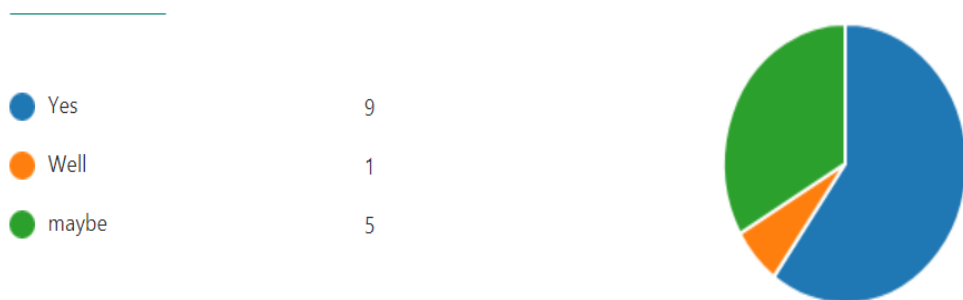
Yes	9
Well	1
maybe	5

Question 5: Would you like to use this application in Future?

Feedback:

🔵	Yes	9
🟠	No	0
🟢	Maybe	6

Question 6: How rapid was the response on search filter?

Feedback:

🔵	Slow	0
🟠	Fast	8
🟢	Very Fast	7

Question 7: Would you like to suggest any changes for future or improvement?

Feedback:

- No its good.
- Add more messages to the content area, e.g when results are not found. And maybe replace the alert windows with something like that. Do not hide table head row when filtering results.
- Maybe if one person has many different expertises, they should only be listed once. Now I can get a filtered list that consists of mainly one person.
- If it is open for members and non members, it would be great.
- gets better when there are more data
- Short experties description.
- Some Fancy look. With professor name the image.
- Communication and less time consuming contacts.

- Commuinication and data information Loop. The process should go on. The publication references should be there.
- Everything looks good.

Every individual user completed the survey independently. All the feedbacks were anonyms. Already some changes have been madebased on the feedbackThe future work is discussed in next chapter.

6 CONCLUSION AND FUTURE WORK

6.1 Conclusion

After taking in to account many aspect into account and study of relevant applications such as Vertical, Newco Helsinki and Arctics the application is finally developed. These relevant systems helped in detail understanding of matchmaking concepts. The application development has already reached on its conclusion point. The application require a lot time and work in data gathering, data manipulation and data arrangements.

The main purpose of this thesis is answering three research questions (RQ). Initially how to create application that matches experts (RQ1)? Secondly, the possible solutions to find an expert inany industry, separation target or separation technology is known (RQ2)? Lastly, how matchmaking application can be helpful for the industry, individual or organizations (RQ3)?

To answer all the research questions a systemic literature review performed. Various systems studied how they worked and how matchmaking algorithms works in different applications. Firstly, all the major data of experts was collected at one place using different sources, as data gathering was a separate challenge. After data gathering process, an interface was generated that helps the user in matching and finding users. Secondly, the data that was collected initially that also consists of information about industry, separation technology and separation target. The information was stored in the MongoDB database, an interface developed for the ease of users to interact with the information. In order to interact and fast response of service, different algorithm generated for matchmaking.

Lastly, after a lot consideration this application was developed to help the industry, users and other people that are connected to it in the separation industry. It matches peoples requirement directly without searching separately on google or digging deep into different locations. All the information is already filtered from various sources so

that the user does not need to visit many web pages for the information or required resource.

6.2 Future Work

There are some aspect that are considered for future work, which requires more time for the development. Using Artificial intelligence is possible in the work, using artificial intelligence will help the application in developing more rapid and accurate matchmaking around the world, it will help for further suggestions of related product or experts using different customer service request accordingly.

7 REFERENCE:

[1] (Nov 01,). *A Brief History of Data Storage.* Available: http://www.dataversity.net/brief-history-data-storage/.

[2] (). *Architecture of PyPI.* Available: https://www.aosabook.org/en/packaging.html.

[3] (). *Process.* Available: https://www.vertical.vc/process/.

[4] (). *The Whole is Greater
Than the Sum of its Parts.* Available: https://www.vertical.vc/about/.

[5] (). *Deal Room.* Available: https://arctic15.com/deal-room/.

[6] (). *Building communities with uncapped opportunities.*
Available: http://arcticstartup.com/about-us/.

[7] J. Mynarz, V. Svátek and T. Di Noia, "Matchmaking Public Procurement Linked Open Data," 2015.
Available: https://www.openaire.eu/search/publication?articleId=dedup_wf_001::27099b33dbb872db790347c76c514dcc. DOI: 10.1007/978-3-319-26148-5_27.

[8] (Dec 01,). *AGILE SOFTWARE DEVELOPMENT VS. WATERFALL SOFTWARE DEVELOPMENT.* Available: https://www.smartsheet.com/understanding-agile-software-development-lifecycle-and-process-workflow.

[9] (Mar 31,). *Python Flask Windows Development Environment Setup.*
Available: http://timmyreilly.azurewebsites.net/python-flask-windows-development-environment-setup/.

[10] (). *Install Flask.* Available: http://flask.pocoo.org/docs/0.12/installation/.

[11] (Apr 30,). *Python Frameworks Used for Application-Oriented Fields.*
Available: https://dzone.com/articles/python-framework-used-for-application-oriented-fie.

[12] (). *Jinja.* Available: http://jinja.pocoo.org/.

[13] Brown *et al*, *The Architecture of Open Source Applications.* (Volume 1 ed.) 2012.

[14] (). *What is Python? Executive Summary.*
Available: https://www.python.org/doc/essays/blurb/.

[15] (Jan 24). *How to set the path and environment variables in Windows.*
Available: https://www.computerhope.com/issues/ch000549.htm.

[16] (2011-). *Plp installtion in windows.* Available: https://docs.python-guide.org/starting/install3/win/#install3-windows.

[17] (May 23). *Installing Python.* Available: https://realpython.com/installing-python/.

[18] (Mar 07). *Environment variable in Windows.*
Available: https://www.digitalcitizen.life/simple-questions-what-are-environment-variables.

[19] Steve Francia, *MongoDB and
PHP.* 2012Available: https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=847593.

[20] (). *MOngoDB Setup*. Available: https://docs.mongodb.com/v3.2/tutorial/install-mongodb-on-windows/.

[21] (). *MongoDB*. Available: https://www.mongodb.com/what-is-mongodb.

[22] ("Oct 18 "). *The Evolution of Database*.
Available: https://mhaadi.wordpress.com/2010/10/18/the-evolution-of-database/.

[23] (NOVEMBER 24,). *Types of database management system and their evolution*.
Available: https://www.analyticsvidhya.com/blog/2014/11/types-databases-evolution/.

[24] K. Driscoll. (). *From Punched Cards to "Big Data": A Social History of Database Populism*.
Available: http://www.dx.doi.org/10.7275/R5B8562P. DOI: 10.7275/R5B8562P.

[25] R. Rudman and R. Bruwer, "Defining Web 3.0: opportunities and challenges," *The Electronic Library,* vol. 34, *(1),* pp. 132-154, 2016.
Available: http://www.emeraldinsight.com/doi/abs/10.1108/EL-08-2014-0140. DOI: 10.1108/EL-08-2014-0140.

[26] (). *"Web 3.0: The Third Generation Web is Coming"*.
Available: https://lifeboat.com/ex/web.3.0.

[27] S. Schulte, "Web Service Discovery Based on Semantic Information - Query Formulation and Adaptive Matchmaking," 2010.

[28] G. Cheng and Y. Qu, "Searching Linked Objects with Falcons: Approach, Implementation and Evaluation," *International Journal on Semantic Web and Information Systems (IJSWIS),* vol. 5, *(3),* pp. 49-70, 2009. Available: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/jswis.2009081903. DOI: 10.4018/jswis.2009081903.

[29] C. Bizer, T. Heath and T. Berners-Lee, "Linked Data - The Story So Far," *International Journal on Semantic Web and Information Systems (IJSWIS),* vol. 5, *(3),* pp. 1-22, 2009.
Available: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/jswis.2009081901. DOI: 10.4018/jswis.2009081901.

[30] S. Sizov, "What Makes You Think That? The Semantic Web's Proof Layer," *Mis,* vol. 22, *(6),* pp. 94-99, 2007. Available: https://ieeexplore.ieee.org/document/4397215. DOI: 10.1109/MIS.2007.120.

[31] N. K. Janjua and F. K. Hussain, "Development of a logic layer in the semantic web: Research issues," in Nov 2010, Available: https://ieeexplore.ieee.org/document/5663555. DOI: 10.1109/SKG.2010.61.

[32] M. Alexander and S. Staab, "Ontology Learning for
the Semantic Web," 2001.

[33] (). *Rdf-schema*. Available: https://www.w3.org/TR/rdf-schema/#ch_introduction.

[34] (). *Resource description framework*. Available: https://www.w3.org/RDF/.

[35] Anonymous "International journal of advanced research in computer science and software engineering,".

[36] T. Berners-Lee and M. Fischetti, "Weaving the web," in *Weaving the Web*Anonymous Harper San Francisco, 1999.

[37] C. Jouis *et al*, "Indexing the World Wide Web: The journey so far," in *Next Generation Search Engines: Advanced Models for Information Retrieval*Anonymous 2012,

Available: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-0330-1.ch001. DOI: 10.4018/978-1-4666-0330-1.ch001.

[38] (). *W3C Semantic Web Activity *. Available: "http://www.w3.org/2001/sw/", World Wide Web Consortium, 2001.

[39] HanXiaotiNniguL *et al*, "SubjIencftormIantieognroafHtiigohEnedrucaItnisotnitiuntCitohnetsoexft Web3.0," .

[40] C. K. Chan, Y. C. Lee and V. Lin, "Harnessing web 2.0 for collaborative learning," in 2009, Available: http://scholarbank.nus.edu.sg/handle/10635/109751.

[41] Anonymous "Web 3.0 in Education & Research," *BVICAM's International Journal of Information Technology,* 2011.
Available: https://www.openaire.eu/search/publication?articleId=doajarticles::5837fa714d7613e0f58e87026b726ca9.

[42] (September 06,). *Is Web 3.0 Really a Thing?*. Available: https://www.lifewire.com/what-is-web-3-0-3486623.

[43] H. Chu and C. Xu, "Web 2.0 and its dimensions in the scholarly world," *Scientometrics,* vol. 80, *(3),* pp. 717-729, 2009. . DOI: 10.1007/s11192-008-2103-y.

[44] Rob Procter *et al*, "Adoption and use of Web 2.0 in scholarly communications," *Philosophical Transactions: Mathematical, Physical and Engineering Sciences,* vol. 368, *(1926),* pp. 4039-4056, 2010.
Available: https://www.jstor.org/stable/25704698. DOI: 10.1098/rsta.2010.0155.

[45] C. Jouis *et al*, "Indexing the World Wide Web: The journey so far," in *Next Generation Search Engines: Advanced Models for Information Retrieval*Anonymous 2012, Available: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-0330-1.ch001. DOI: 10.4018/978-1-4666-0330-1.ch001.

[46] Jian Wu and Zhaohui Wu, "Similarity-based web service matchmaking," in 2005, Available: https://ieeexplore.ieee.org/document/1531265. DOI: 10.1109/SCC.2005.93.

[47] Robert J T Morris and Brian J Truskowski, "The evolution of storage systems," *IBM Systems Journal,* vol. 42, *(2),* pp. 205, 2003.
Available: https://search.proquest.com/docview/222419839.

[48] Robert J T Morris and Brian J Truskowski, "The evolution of storage systems," *IBM Systems Journal,* vol. 42, *(2),* pp. 205, 2003.
Available: https://search.proquest.com/docview/222419839.