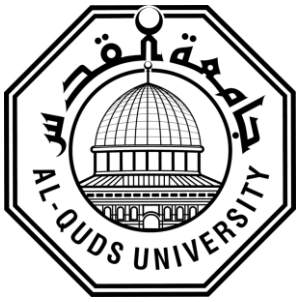
**Deanship of Graduate Studies**
**Al-Quds University**




# Automatic Matching Engine:

# Towards Enhanced Finding of Jobs & Learning Opportunities

**Yousef Mashhour Ibrahim Sabbah**

**M.Sc. Thesis**

**Jerusalem-Palestine**

**1438 / 2016**

# Automatic Matching Engine

# Towards Enhanced Finding of Jobs & Learning Opportunities

**Prepared By:**
**Yousef Mashhour Ibrahim Sabbah**

**B.Sc.: Computer Science - Al-Quds University - Palestine**

**Supervisors: Dr. Jad Najjar**

**A thesis submitted in partial fulfillment of requirements for the Master's Degree in Computer Science / Department of Computer Science / Faculty of Graduate Studies – Al-Quds University.**

**1438 / 2016**

**Al-Quds University**
**Deanship of Graduate Studies**
**Computer Science Department**


## Thesis Approval

# Automatic Matching Engine: Towards Enhanced Finding of Jobs & Learning Opportunities

**Prepared By: Yousef Mashhour Ibrahim Sabbah**

**Registration No: s0912045**

**Supervisors: Dr. Jad Najjar**

Master thesis submitted and accepted on: 04/01/2017
Examining Committee Members (Name & Signature):

1- Head of Committee(Supervisor): Dr. Jad Najjar Signature: .....

2- Internal Examiner: Dr. Badie Sartawi Signature: .....

3- External Examiner: Dr. Ahmed Ewais Signature: .....

Jerusalem – Palestine

**1438 / 2016**

# Dedication

This work is dedicated…

To my parents for their love, endless support and encouragement…

To my beloved wife, without her caring support it would not have been possible…

To my children: MAJD, ALMA and TAIM.

To my brothers, friends and colleagues…

To all of you I say a big

"Thank you" for being an example of love and care.

Yousef Mashhour Ibrahim Sabbah

## Declaration

I certify that this thesis submitted for the Master's Degree, is the result of my own research, except where otherwise acknowledged, and that this study (or any part of it) has not been submitted for a higher degree to any other university or institution.

Signed……..

Yousef Mashhour Ibrahim Sabbah

Date: / /2016

## Acknowledgments

First and foremost Praise be to the Almighty Allah, Lord of all creatures, the Most Gracious and Most Merciful, for all the blessings I have experienced throughout my life.

My sincere thanks for my Supervisors: Dr. Jad Najjar and Dr. Badie Sartawi, for the sincere efforts, interest and time they have spent to guide my research.

I extend my thanks to the members of the Examining Committee: Dr. Jad Najjar, Dr. Rashid Jayousi, Dr. Badie Sartawi, Dr. Nidal Kafri and Dr. Ahmed Ewais

I am very grateful to all professors at the Department of Computer Science in Al-Quds University for their support and dedication.

My thanks are extended to all those who have participated in the evaluation process. Thank you for the time you dedicated.

Finally, and most importantly, I would like to thank my wife. Her support, encouragement, quiet patience and unwavering love are undeniable.

## Abstract

Data about individual skills, knowledge, learning opportunities and job opportunities are distributed over different systems and are based on different schemas and technical formats. These schemas are obtained from different sources such as universities, training institutes, companies, recruitment agencies and the users' CVs and profiles.

This research aims at introducing a new matching engine that provides recommendations for learners in order to select better jobs or courses for the users based on their achieved learning outcomes (skills, knowledge, competence) on one hand, and learning the outcomes required for a job or course on the other, in addition to the management of the interviewing steps to provide recommendations for the most suitable job for each user.

The recommendation system is based on text classification algorithms, especially Naïve Bayes Classification [1]. The idea is to classify texts based on the posterior probability of the documents belonging to the different classes on the basis of the word presence in the documents, and to use Levenshtein distance (LD) [2] as a measure of the similarity between two strings, which we will refer to as the source string (s) and the target string (t).

**محرك بحث لايجاد التطابق المناسب ما بين الوظائف والمقررات الدراسية**

**إعداد: يوسف مشهور ابراهيم صباح .**

**المشرف: دكتور. جهاد النجار**

**ملخص :**

تقوم هذه الدراسة على ايجاد محرك بحث قادر على ايجاد مدى التطابق والتقارب بنسبة مئوية ما بين المقررات الدراسية المطروحة من قبل الجامعات والمؤسسات التعليمة المختلفة من جهة والوظائف المطروحة من قبل الشركات ومؤسسات التوظيف من جهة اخرى على الرغم من تواجد هذه الجامعات وشركات التوظيف باماكن مختلفة وعلى بيئة عمل مختلفة ايضا, تم تطبيق هذه الدارسة بالاعتماد على مخرجات التعليم وميزات وخصائص الوظائف المختلفة من مهارات وخبرات لدى الافراد الراغبين بالتقدم لهذه الفرص المختلفة على اختلافها , ومن ثم العمل على متابعة هذا التطابق لترتيب المقابلات المناسبة لهذه الفئة التي تم اختيارها بناءا على مؤشرات ذات معاييروقيم يتم تحديدها من قبل . حيث تقوم هذه الدراسة على اعطاء التوصيات المناسبة لاصحاب العلاقة لاختيار الشخص المناسب ضمن هذه المعايير .

تم اجراء هذا التطابق بالاعتماد على خوارزمية Naïve Bayes Classification بعد اجراء بعض التعديلات والمعالجات على قيم التطابق المختلفة والمخرجات الخاصة لكل من المقررات الدارسية والوظائف بالاعتماد على خوارزمية Levenshtein distance .

لقراءة هذا المخرجات المختلفة من المصارد التعليمية على أختلافها تم الاعتماد على استخدام نموذجيين رئيسيين , الاول هو Metadata learning Opportunity وهذا خاص بالمقررات التعليمية وخصائصها والاخر هو Personal Achieved learning outcomes وهذا المعيار خاص بايجاد وقراءة مالدى المتعلم من خبرات مهارات وما يطمح للوصول اليه ايضا .

تم التحقق من نتائج هذه الدراسة عن طريق بناء نظام يقوم بتجميع هذه المخرجات على اختلاف مصادرها والاعتماد على المعايير السابقة وباستخدام الخوارزميات المذكورة اعلاه , تم التحقق من النتائج من خلال شريحتين من المتقدمين الاولى : تم تجميع مجموعة من الاشخاص ذوي الكفاءة ومن لديهم مؤهلات معينة وتم تطبيق هذه الدراسة عليهم والفئة الثانية عبارة عن شخصين الاول هو خبير في المقابلات والاخرهو خبير في التوظيف وتبين ان النتائج دقيقة لكلا الفئتين.

## List of Abbreviations

| | |
|---|---|
| **MLO** | Metadata for Learning Opportunities |
| **LO** | Learning Opportunity |
| **LOD** | Learning outcome definition |
| **LOP** | Learning Opportunity Provider |
| **LOS** | Learning Opportunity Specification |
| **LOI** | Learning Opportunity Instance |
| **ALO** | Achieved Learning Outcomes |
| **PALO** | Personal Achieved Learning Outcomes |
| **AR** | Assessment Records |
| **RLO** | Related Learning Outcomes |
| **ILO** | Intended Learning Outcomes |
| **BoK** | A Body of Knowledge |
| **KU** | knowledge unit |
| **KA** | Knowledge area |

# Table of Contents

## List of Tables

# List of Figure

## Chapter 1



## Chapter 2

**Chapter 3**

**Chapter 4**



**Chapter 5**

## Keywords

Learning outcomes, assessment, learning opportunities, learning needs, personal achievements, job profiles, structured data, Unstructured data, Naïve Bayes Classification, semantic search, edit distance (Levenshtein's), Stemming.

# Chapter 1

## 1. Introduction

### 1.1. **Overview**

When students graduate from universities, educational centers, training centers or complete special courses or programs, the first thing they start doing is searching for jobs that meet their needs and aspirations. Some students join new courses and additional educational programs for more development and to improve their skills, knowledge and competence in the same field of specialization in order to attain greater chances in competing for a job as compared with others job seekers.

The job seekers start searching for jobs in many different ways: they start exploring all available resources like magazines, newspapers, recruitment websites, and social networking sites, other job seekers personally go to the headquarters of companies or institutions to present employment applications, and others subscribe through SMS services that send job announcements using text messages.

When a job seeker finally gets an interview for a job in an organization, there are several other stages that need to be completed in order to get this job. In case the first step was completed successfully, the job seeker will then move to the next stage; which is conducting either a practical or theoretical exam to measure certain characteristics and skills required for the job, and upon passing the related exam the job seeker will conduct personal interview with the prospect employer in order to explore the character of the job applicant, and to identify the points of strengths and weaknesses. Some features are

measured based on experience or based on previously prepared forms that focus on the measurement of specific points, each point is assigned a specific weight in the total score that measures the job seeker's capabilities and achievements, which will be compared with other applicants.

Job Seekers or learners can apply for any jobs or courses regardless of place of residence and employment. Jobs and courses became open to all regardless of distance and time to exchange of resources and jobs, where education has become for all.

Information exchange has become one of the most important foundations of the success for many of communities and institutions of different work; there are no borders, languages of candidate, and different working environments.

In the past, the process of information exchange was performed manually using models and paper, which is a time and effort consuming process, in addition to being susceptible to many mistakes and problems, either during the introduction of data or the validity and suitability and credibility, or the credibility of the source. With the advancement of technology, the internet and other quick methods of data transfer, the process of information exchange became less time-consuming and more confidential. Information became easier to access, enabling educational institutions to view the educational opportunities for students of different specialties, their knowledge, their experience, their languages and their places of residence through their electronic devices, or employment institutions could announce their need to impose functional through their websites as well, or through brokers who display these jobs and identify people to go right to these jobs traditionally, revolutionize the appearance of many websites that provide employment and communication between the student and the companies or institutions, such as the service site LinkedIn, so that the person Create his Profile that contains personal information, qualifications, experience, training courses and other Information that is necessary to get a

job. On the other side, we have the Companies who need Job Opportunities by creating Company Profile and announce the Job qualifications on their profile, with the required qualifications and other Requirements.

Despite the importance of the existence of such high speed data transfer methods, there are, however, symptoms of many problems, including:

1. Overlooking some information when creating the students' files.
2. Incorrect data entry, whether this act is intended or unintended.
3. Uncertainty about the accuracy of the data entered by institutions or recruitment firms.
4. The lack of scientific foundations and models for data transfer, the only method is filling a special form to do so.
5. High time and efforts consumption.
6. Vulnerability to penetration, data loss, and closure of necessary accounts.
7. The inability to find an educational path clear to the learner about what is appropriate and has been recommended to him.
8. Does not provide a career path for the person.
9. Manual modification of personal information, experience and skills.
10. There is no relationship between the learner and the interviews in such sites.
11. No management of interview for a person who has a job opportunity.
12. No provision course opportunity.
13. Specific deployment.
14. No centralization and data management.

## 1.2. **Problem Statement**

Employers seek people who possess particular qualifications, and graduates seek jobs that match their qualifications. The interoperability, manageability and re-usability issues are the main challenges of communities and systems that deal with learning outcome information.

Different communities and systems may use different data models to represent information on skills, knowledge or competencies.

Learning opportunities are described based on different schemas and sources. Job opportunities are also based on different schemas and sources, and the user profile based on different schemas and source. Currently, these three objects are developed separately without integration between them, the information for each one is not managed and centralized. All information obtained from these components is very useful and important because it comes directly from the source and this information needs to be managed and centralized to become more useful.

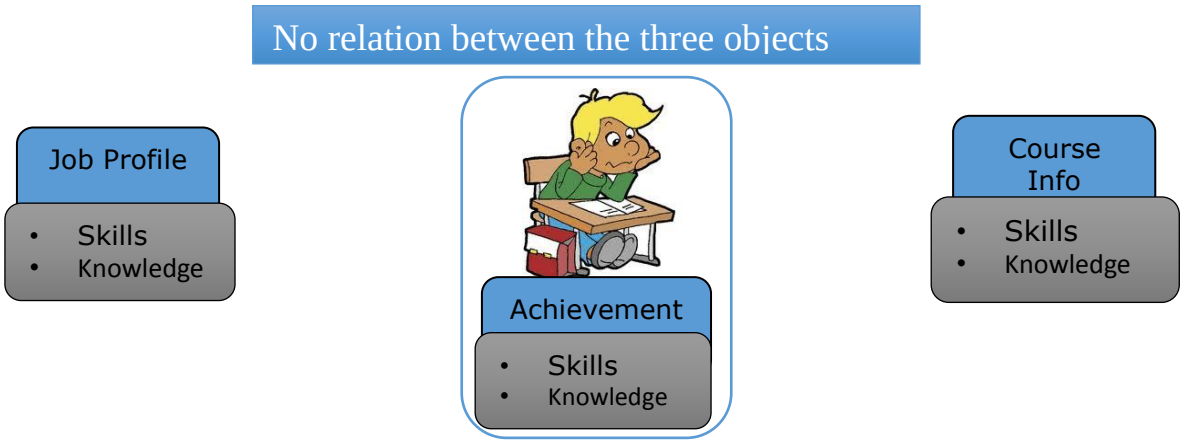



Figure 1:1 : Different Schemas and Sources

## 1.3. **Objectives**

The objective of this research is to introduce a recommendation system that can provide recommendations for the learner between jobs and courses, and to:

1. Find a solution that can exchange data from different environments and sources.
2. Find a solution that facilitates the exchange of data for learning outcomes of the learning opportunity, job opportunity and users achieved learning outcome for students or employee between many universities or educational institutions at different places (learning management systems, e-portfolios, social applications and recruitment systems).
3. Provide recommended jobs or courses depending on personal skills, knowledge, experience and attitudes.
4. Enable the learner to move between universities for admission to their programs, specializations, without the need to exchange traditional papers or certificates.
5. Facilitate the process of finding a historical profile containing the student's educational path or career path for employees, which contains the intended learning outcomes (knowledge, skill and competence) of learning opportunities.
6. Manage interviews for job opportunities to provider and manage courses for opportunity providers.
7. Track the achievements, and related evidence records, of learners after the successful completion of learning opportunities, and required learning outcomes and qualifications.

## 1.4. **The Solution**

In this thesis, we will develop an **automatic matching engine** that uses personal achieved learning outcomes and compares them to the learning opportunities outcome for jobs and learning outcomes for course opportunities to get the recommendation for what the persons can have (Job, course), or which path they can follow, and manage the interview process to select the most suitable person for specific jobs.

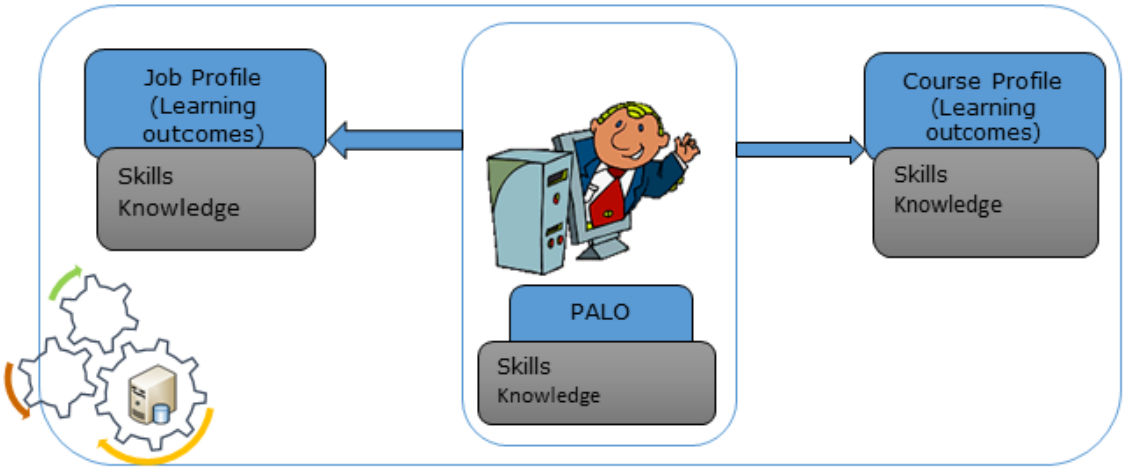



Figure 11:2: Matching Engine Item Relations

## 1.5. **Methodology**

This section provides information about the approach used for conducting this research. The following steps were followed in the development of the matching engine and algorithms that can calculate the percentage of matching between a learning opportunity and job opportunities.

- **Data Collection**: Several data collection methods were used for this research. The first methods is interviews; we conducted interviews to collect data from decision makers in order to be able to select the best candidate based on the archived learning outcomes and we collected the data about the course from institutes and training centers using the following methods:

1. We held many meetings with our Supervisors on the campus of the Al-Quds University, and other meetings through Skype to discuss the idea and to set the main points.
2. We held meetings with experts from Italy and Turkey to discuss the mechanism linking universities in Turkey manner model **MLO**.
3. We visited local educational institutions and companies, especially training centers, and met with many of the graduates and students who joined training courses and asked them many questions :
    - Would you prefer to search for a specific course on the Internet?
    - Would you like to join the course?
    - Do you prefer to search for a course through the Internet, magazines or through social networking sites?
    - What do you think about allowing universities or educational institutions to publish your learning information through the Internet and allow others to review them?
    - How can you get an idea about the size of the market demand for your qualification?
    - How can you improve your experience, knowledge and skills?
    - Do you prefer to have a system that can enable the introduction to your skills, knowledge, and experience and propose what's appropriate for you?
    - Did you interview for some jobs?
    - What are the interview stages that you went through, and which ones did you pass?
    - Do you believe that the interview process is fair?

- Do you believe the interview process presents what you have exactly?
- Did you reach the last stage for any job Interviews?
- How many interviews have you done?
- How many training courses have you got? Why?
- What are the courses that you require?
- During the interview, did the interviewer ask you for some information that was not stated in your resume, or out of your skills and knowledge?

4. We held a meeting with employee groups in our local market, and asked them some questions such as :
    - Do you need any training courses?
    - Why do you need training courses?
    - Do you believe the course improved your skills and knowledge?
    - Do you prefer to get another Job? Why?
    - How do you stay updated about the latest technologies in your field of specialization?
    - Do you prefer to find a solution that can recommend suitable course or job depending on your experiences (skill, knowledge)?
    - Do you prefer to get a training course from a well-known educational institution?
    - Do you prefer to find a solution that can evaluate your skills and knowledge and compare them to the skills and knowledge of other people?
    - Do you prefer a solution that gives you recommendation for jobs or training centers?
    - Do you prefer to find solution that can give you advice about your career path or learning path?

- Do you prefer to find a solution to give fair chance between all learners or job seekers?

5. We held some meetings with representatives of the human resource departments in some local companies, such as Hadara Information Technologies, Jawwal Cellular Co., and Paltel Company, and asked them several questions, including:
    a. What is the form used to request a new employee?
    b. What are the main items that you focus on in a job application?
    c. What are the basis of interviewing?
    d. What are the basis of differentiating between competitors interviewing for a particular job?
    e. What are the parameters that you use to select someone?
    f. Do you prefer that there be a system capable of proposing appropriate employees to you based on the job application?
    g. Do you prefer to obtain the largest number of applicants?
    h. Do you need to have a job profile for your company?
6. We have opportunity to apply the system in Jawwal Company, especially on program called (Go Professional) that Jawwal uses to get employees every year from different universities for working in Jawwal. They made interviews in universities to employ a person who has skills and Knowledge in some specified fields.
7. We held meetings with a local recruitment company (jobs.ps company) and asked some questions, including the following:
    a. Do you have websites that present job descriptions?
    b. How do you collect the job seekers resumes?
    c. How do you manage the interview process?

d. Do you prefer to find a solution that manages all candidates' processes?

The other methods for data collection are:

1. Searching the web for more information regarding this topic. Through this search, I was able to get many results, which increased my knowledge in this subject, such like Google, YouTube, LinkedIn, download and many research papers, workshops studies and analysis related to this topic.
2. Finding existing projects related to my research topic, such as:
   - http://www.icoper.org
   - http://www.medbiq.org/node/819
   - http://www.openscout.net/
   - http://wiki.teria.no/display/inloc/Home
   - http://www.prolixproject.org/

After collecting all requested data, the next step was to analyze this data in order to find the background of the solution with input, process and output.

In Chapter 2, Section 2.2 of this Thesis, you can find more details regarding this point.

- **Background and Standards**

In this section we will introduce the standards and algorithm that we used to approve and develop the engine.

1. **Learning Outcome Definitions (LOD)** is a data model that defines a conceptual base schema for describing and sharing learning outcome definitions in the context of online and technology enhanced learning.
2. **Metadata for Learning Opportunities (MLO)**, MLO is a European

specification that covers a wide range of information about learning opportunities including intended learning outcomes with the following items:

1. Learning Opportunity Provider: An agent (person or organization) that provides learning opportunities.
2. Learning Opportunity Specification: An abstract description of a learning opportunity, consisting of information that will be consistent across multiple instances of the learning opportunity.
3. Learning Opportunity Instance: A single occurrence of a learning opportunity.
4. Personal Achieved Learning Outcomes (PALO) is a simple schema proposed to capture information on knowledge, skills and competences achieved by a learner, and the relations between those outcomes, represent information on achieved/required/desired learning outcomes of a learner, also teachers (taught outcomes) and share data with learning management Systems, recruitment systems, and social applications.

3. **ACM/IEEE** the ACM and IEEE Computer Society jointly sponsor the development of a Computing Curricula volume on Computer Science. These volumes have helped to set international curricular guidelines for undergraduate programs in computing. They aimed at providing modern curricular guidance for undergraduate Computer Science programs internationally, It's a standard that categorize a set of item with the same category.

3. **Levenshtein Distance Algorithm**, is a measure of the similarity between two strings, which we refer to as the source string (s) and the target string (t). The Levenshtein edit distance is the number of insertions, deletions, or replacements of single characters that are required to convert one string to the other.

4. **Bayesian Classifiers**: In Bayesian classifiers (also called generative classifiers), we attempt to build a probabilistic classifier based on modeling the underlying word features in different classes. The idea is then to classify text based on the posterior probability of the documents belonging to the different classes on the basis of the word presence in the documents. The Naive Bayesian technique is the most popular and is used in contemporary spam filtration, document categorization, and help desk systems because of its accuracy, robustness, and ease of implementation

- **Evaluation and Testing**

In this section, we analyze the collected data and the system standards using the algorithms to test and evaluate the results and the performance of the system. In order to achieve this, we have conducted manual experiments with real data and applied the same example in the developed system to compare the result. We have collected real data for achieved learning outcomes from five participants who are employees at Hadara Technologies Company along with their profiles, and three jobs skills and descriptions from the recruitment company (Jobs.ps). However, the course skilled and knowledge data collected from Expert turnkey solution – Ramallah and Palestine National Institute for Information Technology NIIT - Ramallah to apply our study methodology and for testing our approach we select 2 Expert candidate for make the testing the first one is Job Company Manager and the second one is the head section of HR in Hadara Technology. More details can be found in Chapter 5 of this Thesis.

## 2. Background and Standards

In this section, we will demonstrate the standards we used to import and read the metadata learning opportunity, personal achieved learning profile and job profile.

### 2.1. **Learning Outcomes**

Learning outcomes are specific statements of what learners will be able to do (action verb) under what conditions (by the end of the course).

Importance of learning outcomes:

1. Tell students what they should be able to do at the end of the course,
2. Outcomes are the basis for delivery of content, activities, assessments…etc.
3. Guide the planning of activities and assessments that facilitate the accomplishment of the outcomes.

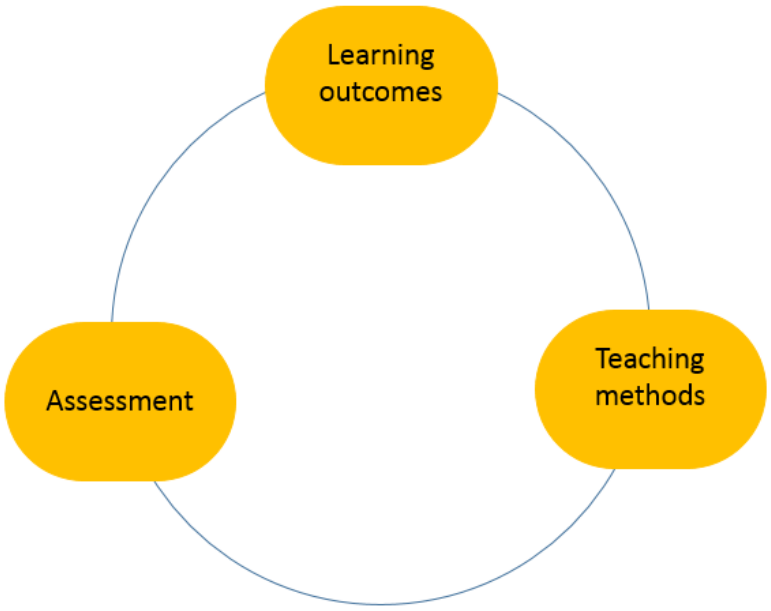



Figure 1:3 Learning Outcome Cycles [3]

The **assessment defined** as a description of the broad approach to assessment is used in the learning opportunity. We have two types of role assessments, the first one is formal assessment, which occurs when the assessment is an integral part of the learning activity. The other assessment is informal assessment, which occurs when the assessment is not part

of the learning activity. Formal assessment produces formal result according to formally defined process (pass/fail) or marks, it means that learners have evidence from the institution that has offered the learning opportunity.

**Teaching method**, which is a generic description of a set of learning outcome oriented teaching and learning activities, may be used as a learning design resource to guide the design of learning opportunity instances.

**Unit of learning**, which represents a complete, self-contained unit of education or training, can be used to design and run learning opportunity instances.

Types of learning outcomes:

1. Knowledge: what the person knows, e.g. the person learns and knows the concepts of the programming language.
2. Skills: the person person's ability to apply the knowledge. E.g. the use of programming langue concepts to solve the problems and find the suitable solution and algorithm.
3. Competence: are more complex; they involve the application of knowledge and skill.

### 2.2. **Learning Outcomes Defination (LOD)**

The Learning Outcome Definitions (LOD) data model defines a conceptual base schema for describing and sharing learning outcome definitions in the context of online and technology enhanced learning.

The data model provides a way to capture the key characteristics of a learning outcome, independently of its use in any particular context or target group (persons).

This model should enable the storage, findability and exchange of learning outcomes across learning systems that deal with learning outcomes data.

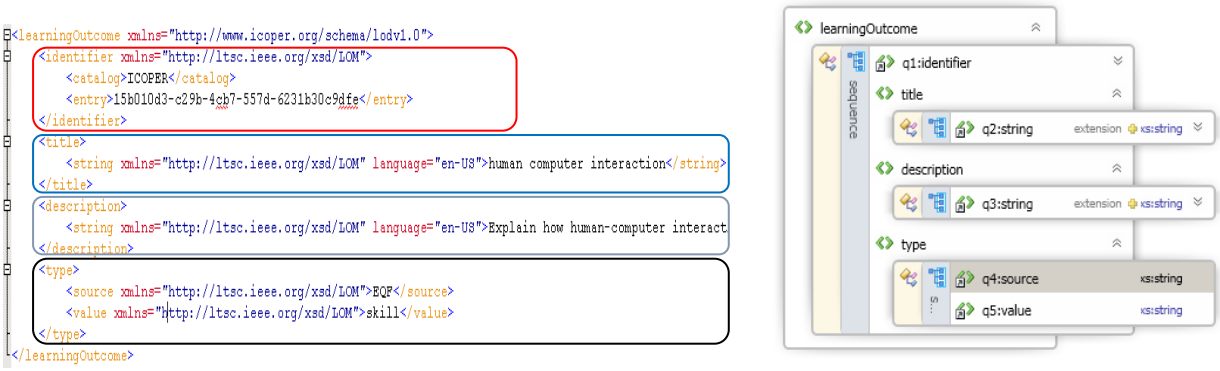



Figure 1:4 Representation Learning Outcomes Definition Schema Using Visual Studio

**Case study** :

Yousef is a teacher at Al-Quds University, he teaches Human-Computer Interaction course for second year students. He specifies Intended learning outcomes for this course and what the students will achieve when they finish this course, and what they will be able to apply when they complete and pass the exams, activities and assessments.

Majd is a student in the same university and he is in his second year at the Computer Science Department. He joined the Human-Computer Interaction course. When he completes the course, he aspires to be able to apply the concepts of this course to his career and can build prototypes and scenarios, and he needs to apply these course concepts for some specified jobs.

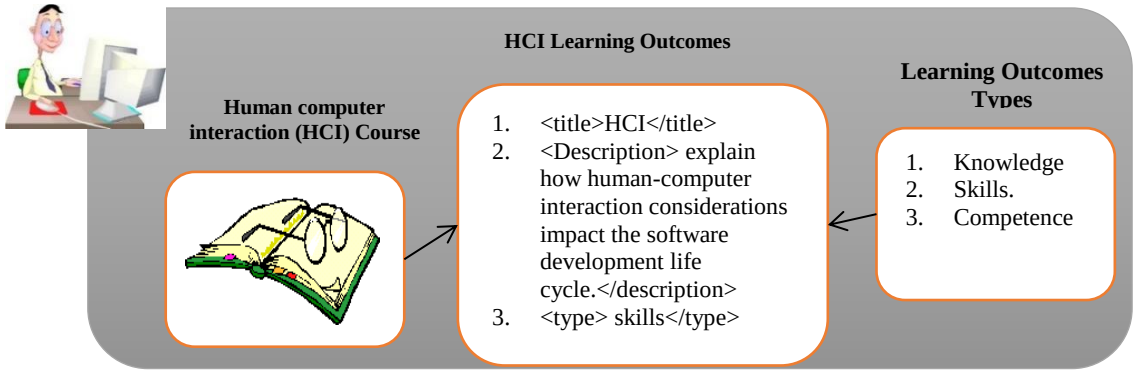



Figure 1:5 Representation of Learning outcomes Definition Process, e.g: HCI Course Learning Outcomes

## 2.3. **Metadata for Learning Opportunities (MLO)**

MLO is a European specification that covers a wide range of information about learning opportunities including intended learning outcomes.

The key concepts are defined as follows:

- Learning Opportunity Provider: An agent (person or organization) that provides learning opportunities.
- Learning Opportunity Specification: An abstract description of a learning opportunity, consisting of information that will be consistent across multiple instances of the learning opportunity.
- Learning Opportunity Instance: A single occurrence of a learning opportunity.

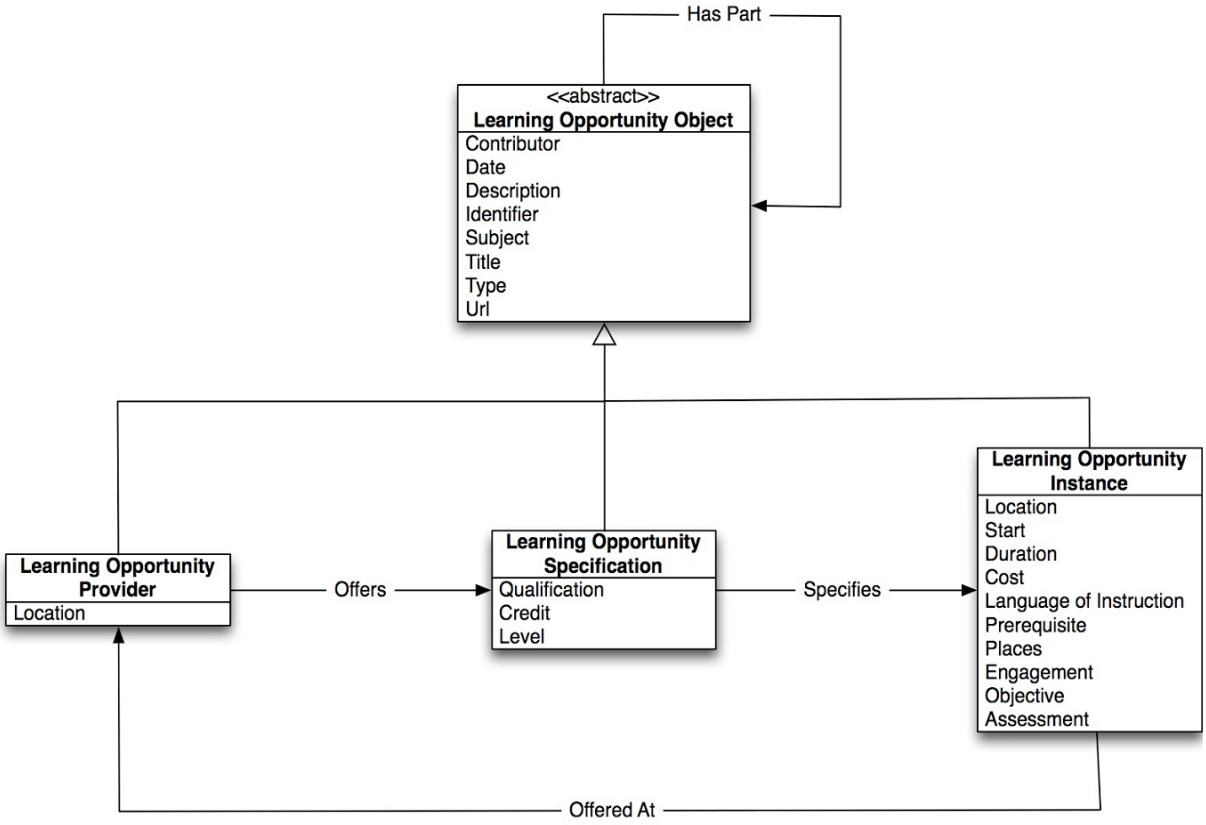



Figure 1:6 MLO Elements and Relations (UML diagram) [4]

MLO records information about intended learning outcomes of an opportunity in free text format, the interoperability and reuse of common learning outcome definitions across learning opportunities and curricula, some learning outcomes may be common between several learning opportunities of the same program or across universities.

Learning outcomes should be managed independently of learning opportunity metadata and be linked using a semantic web services; e.g., using a URI.

**The MLO objectives** element can capture multiple learning outcomes for each learning opportunity instance that can be referenced using a URI.

**Metadata for Learning Opportunities (MLO) Elements**


Figure 1:7 Representation Metadata Learning Opportunities Elements and Relations, using visual studio class Diagram

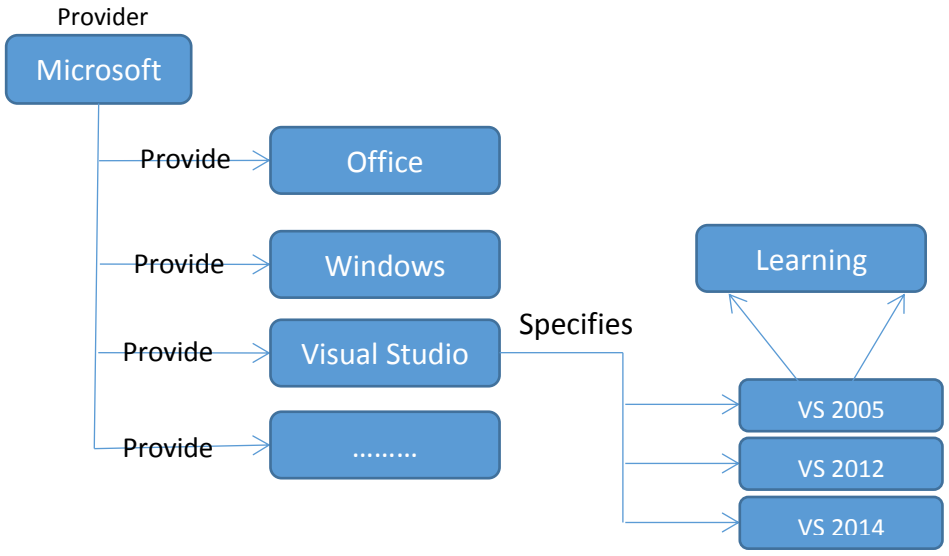



Figure 1:8 Examples for the Hierarchical of the MLO Items. Microsoft has many products like windows and this category has many versions like xp, win server, win7 or win10

| **Item Name** | **Description** |
|---|---|
| **Learning Opportunity Provider** | The provider who provides the learning opportunity, such as universities |
| **1.1.** Identifier | Primary identifier for provider |
| **1.2.** Subject | |
| **1.3.** Title | |
| **1.4.** url | |
| **1.5.** location | |
| **2. learning Opportunity Specification** | |
| **2.1.** identifier | |
| **2.2.** subject | |
| **2.3.** title | |
| **2.4.** url | |
| **3. learning Opportunity Instance** | The learning opportunity attributes (Course) |
| **3.1.** contributor | The contributor of the course |
| **3.2.** description | Course description |
| **3.3.** identifier | Primary Identifier |
| **3.4.** title | Course title |

| | |
|---|---|
| **3.5.** url | Course url form the source |
| **3.6.** location | The course will held location |
| **3.7.** start | The start date of the course |
| **3.8.** duration | The period of the course will taking |
| **3.9.** language_of_instruction | The main langue of the material of the course |
| **3.10.** places | |
| 1.1. engagement | |
| 1.2. objective | has sub attributes |
| 1.3. assessment | Has sub attributes |

Table 1: Summary the MLO Attributes

### 2.4. **Personal Achieved Learning Outcomes (PALO)**

The Personal Achieved Learning Outcomes (PALO) data model is a simple schema proposed to capture information on knowledge, skills and competences achieved by a learner, and the relations between those outcomes, represents information on achieved/required/desired learning outcomes of a learner, also teachers (taught outcomes) and share data with learning management Systems, recruitment systems, and social applications.

PALO enables capturing the following information:

1. Relations between achieved learning outcomes, regardless of the taxonomies or anthologies they belong to.
2. Contextual information on where the achieved learning outcome is obtained or applied;
3. Information about all types of evidence and assessment that prove the achievement of a learning outcome;

4. Information about levels and ranking of an achieved learning outcome, such as proficiency level

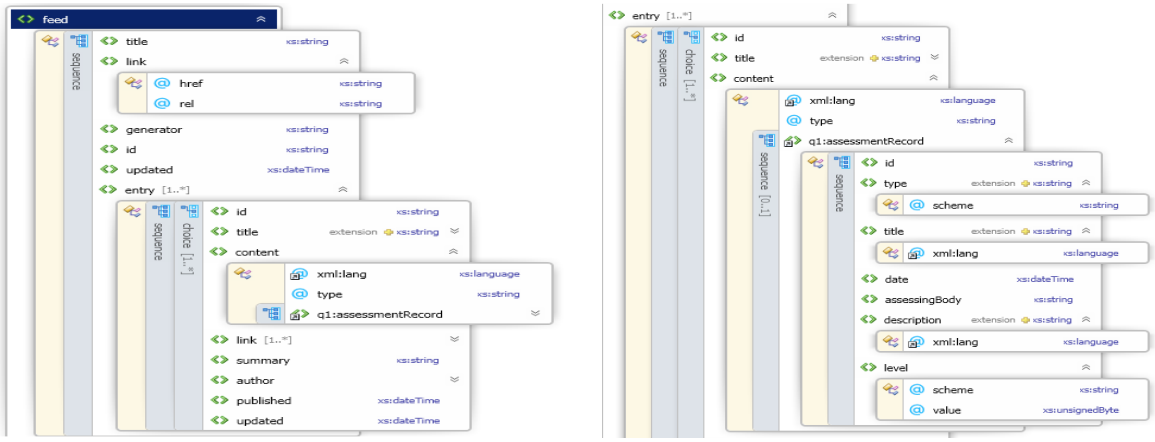



Figure 1:9 Representation PALO Elements and Relations Using Visual Studio

| No | Name | Description |
|---|---|---|
| 1. | **Personal Achievement Profile** | This element represents a collection of one's personal achievements |
| 1.1 | Holder | A human readable title of the personal profile. |
| 1.2 | Identifier | Primary URI that is used to access the profile. |
| 1.3 | Description | A human readable description of the learner profile. |
| 1.4 | Achievements | This element is a reference to the achievements represented in this profile. |
| 2. | **Achievement** | An achievement record, normally of an attained learning outcome. Information about the achievement may be taken directly from a related intended learning Outcome, rather than being given particularly. Personalized versions of title and description may be used to supplement intended learning outcome. |
| 2.1 | Identifier | A globally unique label that identifies the achievements in the individual profile. |
| 2.2 | Title | A text label for the achievement. |
| 2.3 | Description | A human readable description of the achievement. This is a personalized text of the intended learning outcome definition, but can also be similar to intended learning outcome; defined by taxonomy of learning outcomes. |
| 2.4 | Related Learning | Identifier of the intended learning outcome that this |

| | | |
|---|---|---|
| | Outcome | achievement claims to have attained. |
| 2.5 | Context | Identifier of the context where the achievement is claimed to be attained. |
| 2.5.1 | Scheme | A reference to the definition or the schema used to describe the context values. |
| 2.5.2 | Value | A label value/term of the context. |
| 2.6 | Assessment Record | Identifier of the assessment record that stands as evidence to the achievement. |

Table 1 : Summary the PALO Attributes

| **No** | **Name** | **Description** |
|---|---|---|
| 3. | Intended Learning Outcome | The intended learning outcome (knowledge, skill and competence) that is attained by the learner. |
| 3.1 | Identifier | A globally unique label that identifies the learning outcome (Knowledge, Skill and Competence).<br>The Identifier is sufficient to reference the learning outcome definition in any other system. |
| 3.2 | Title | A single mandatory text label for the learning outcome. This is a short human readable name for the learning outcome.<br>The Title may be repeated in multiple languages. Each translation is represented by an instantiation of Lang String type.<br>The identifier provides the definitive reference to the learning outcome.<br>The title element provides a convenient, alternative, readable form. |
| 3.3 | Description | A human readable description of the learning outcome. This is an optional unstructured (opaque) "text blob" meant to be interpretable only by humans.<br>The Description may be repeated in multiple languages. |
| 3.4 | Type | An element that captures the type of learning outcome, according to the European Qualification Framework (EQF). |
| 3.5 | Related Learning Outcome | Captures information of other learning outcomes that may be related to the current described learning outcome. |
| 3.5.1 | Reference | A globally unique label that identifies the intended learning outcome (knowledge, skill or competence). |
| 3.5.2 | Relationship Type | This element captures the type of relation between |

| | | |
|---|---|---|
| | | the current described Learning outcome and another learning outcome.<br>Examples are the SKOS relations: http://www.w3.org/TR/skos-reference/ - like Narrower, broader. |
| 3.6 | Context | Identifier to the context of the intended learning outcome. |
| 3.6.1 | Scheme | A reference to the definition or the schema used to describe the context values |

**Table 1.2 B: Summry the PALO Attributes**

| No | Name | Description |
|---|---|---|
| 3.6.2 | Value | Represents a numeric value for the level. |
| 4. | Level | A set of metadata elements that capture ranking information about the learning outcomes of learners. This includes proficiency level, interest level, weight and ageing. |
| 4.1 | Name | Capture the genre/name of the ranking. It can capture the proficiency level of the learning outcome, the learner interest in obtaining the outcome or the ageing of the outcome. Some learning outcomes may degrade by time, like language skills. |
| 4.2 | Value | Captures a numeric value for the level.<br>Example: the eight EQF levels of proficiency. |
| 4.3 | Scheme | A reference to the definition or the schema used to describe the qualifier level values. |
| 4.4 | Description | A textual description about the level described. This element is useful to be provided when the scheme element is not provided. In other words, when a level value provided is not part of a common ontology or taxonomy. |
| 5 | Context | A set of factors that are external to and give meaning to a learning outcome. For instance subject domain and location (e.g., lab, classroom) are textual information that gives meaning to the learning outcomes. |
| 5.1 | Value | A label value/term of the context. |
| 5.2 | Scheme | A reference to the definition or the schema used to describe the context values. |
| 5.3 | Description | A textual description about the context domain. This element is useful to be provided when the scheme |

| | | |
|---|---|---|
| | | element is not provided. In other words, when a context value/term provided is not part of a common ontology or taxonomy. |

Table 1.2 C: Summary the PALO Attributes

| No | Name | Description |
|---|---|---|
| 6. | Assessment Record | Captures information of evidence that a learner has obtained a learning outcome. This record constitutes of evidence of the verification of the attainment of a certain achieved learning outcome by a certain learner. Thus, assessment records allow to associate learners and learning outcomes, in a formalized way.<br>Apart from the learner data and learning outcome data, an assessment record<br>Provides information about the type of test performed for verifying the achieved learning outcome, also the responsible expert or institution who endorses it. |
| 6.1 | Identifier | A globally unique identifies to the assessment record The Identifier is sufficient to reference the assessment record in any other system. |
| 6.2 | Type | Captures the form of evidence accepted as a formal proof of the attainment of a learning outcome. |
| 6.3 | Title | Provides a readable description of the assessment record. |
| 6.4 | Score | A numeric value that represents a result/grade of an assessment of certification |
| 6.5 | Date | The date the evidence record is created. |
| 6.6 | Assessing Body | The name or a reference to the agent (e.g. university) that verifies the Assessment record. vCard, as defined by IMC vCard 3.0 (RFC 2425, RFC 2426) |
| 6.7 | Description | A textual description about the assessment value. This element captures the feedback type of |

| | | |
|---|---|---|
| | | assessment where no scores for the assessment are provided. |
| 6.8 | Attached Reference | A reference to any attachments that prove the obtaining of learning outcome and the evidence record |
| 6.9 | Level of Assessment Result | Identifier to the definition or the schema used to describe the level values. |
| 6.9.1 | Scheme | A reference to the schema used to describe the level values. |
| 6.9.2 | Value | Represents a numeric value for the level. |

Table 2 : Summary of the PALO Attributes

The data collected in a person's PALO profile can be used in different ways:

1. It can be used in personal specifications in the learning opportunity of recruitment.
2. Individuals can claim to have attained them.
3. Evidence can be assembled by or about individuals to support a claim to their attainment.
4. They can be used by employers or professional bodies as the basis to review processes that tie in with career progression.
5. This data can also be used for the recommendation of relevant learning opportunities for the learner based on his achievements.

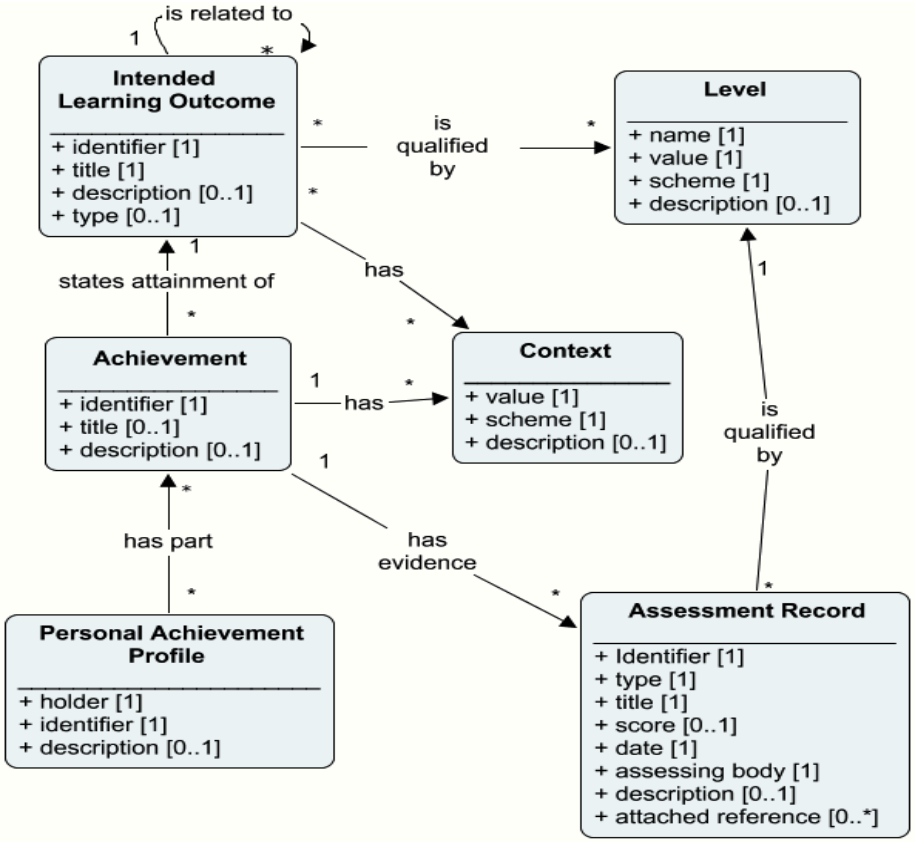



Figure 1:10 PALO Elements and Relations with others [5]

### 2.5. **ACM/IEEE**

The ACM and IEEE Computer Society jointly sponsor the development of a Computing Curricula volume on Computer Science. These volumes have helped to set international curricular guidelines for undergraduate programs in computing. They aim at providing modern curricular guidance for undergraduate Computer Science programs internationally,

It's a standard that categorizes a set of item with the same category.

**A Body of Knowledge (BoK)** specifies a structure of knowledge, skills and learning outcomes that a program must provide. The required knowledge is structured in three hierarchical levels:

1. Knowledge areas (KA).
2. Knowledge units (KU)

3. Topics (in one or two tiers, for core and optional topics), as shown in Fig. 1.

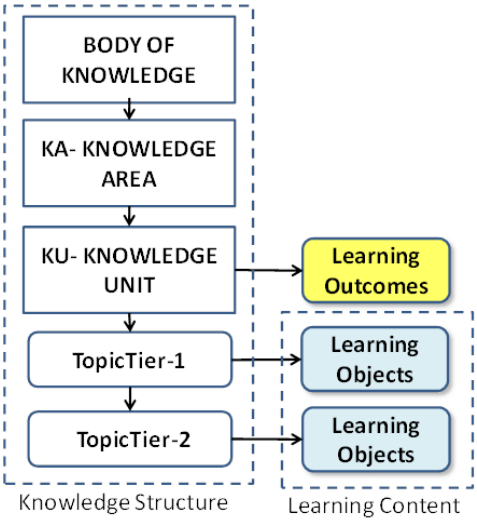



Figure 1:11 The Structure of a Body of Knowledge [6]

Body of Knowledge is presented as a set of Knowledge Areas (KAs), organized on topical themes rather than by course boundaries. Each KA is further organized into a set of Knowledge Units (KUs) as shown in above table,

| | |
|---|---|
| AL - Algorithms and Complexity | NC - Networking and Communications |
| AR - Architecture and Organization | OS - Operating Systems |
| CN - Computational Science | PBD - Platform-based Development |
| DS - Discrete Structures | PD - Parallel and Distributed Computing |
| GV - Graphics and Visual Computing | PL - Programming Languages |
| HC - Human-Computer Interaction | SDF - Software Development |
| IAS - Information Assurance and Security | Fundamentals |
| IM - Information Management | SE - Software Engineering |
| IS - Intelligent Systems | SF - System Fundamentals |
| | SP - Social and Professional Issues |

Table 4: Set of Knowledge Units

A knowledge unit (KU) consists of a set of topics that describe a more complex concept or a set of interrelated concepts that specify a chunk of knowledge that provides students with one or more basic professional capabilities. These are learning outcomes, specifies what students are capable to do, in a professional context, when they learn and acquire knowledge units.

SE/Software Processes

*[2 Core-Tier1 hour, 1 Core-Tier2 hours]*

*Topics:*

[Core-Tier1]

- Systems level considerations, i.e., the interaction of software with its intended environment
- Introduction to software process models.
- Phases of software life-cycles
- Programming in the large vs. individual programming

[Core-Tier2]

- Applying software process models

[Elective]

- Software quality concepts
- Process improvement
- Software process capability maturity models
- Software process measurements

Table 3 : Knowledge Unit Example

Guidance provided on depth of coverage for learning outcomes in each Knowledge Area.

• There are three levels of depth: familiarity, usage, and assessment.

1. **Familiarity**: to know what it means

2. **Usage**: to be able to apply concept (e.g., write the code to use it)
3. **Assessment**: to be able to compare/contrast/select appropriate method/strategy for different situations.

The topics within the KAs will be organized into courses in different ways at different institutions.

Three-tiered classification of Body of Knowledge Units

1. Core-Tier1: essential topics, all of which are required for any undergraduate CS program
2. Core-Tier2: important foundational topics, the vast majority (no less than 80%) of which should be in a CS program
    - Still considered "Core" topics – ideally all Tier2 topics would be included in an undergraduate program, if possible
    - Tier allows for flexibility to locally customize curricula
3. Elective: additional topics that can be included to complete an undergraduate CS program Covering just "core" material is insufficient for a complete curriculum.

| Prerequisites |
| --- |
| Algorithmic Strategies |
| Fundamental Data Structures and Algorithms |
| Interfacing and communication |

**Objectives(learning outcomes)**

| level | Objective |
| --- | --- |
| Familiarity | Explain what is meant by "best", "expected", and "worst" case behavior of an algorithm. |
| Assessment | In the context of specific algorithms, identify the characteristics of data and/or other conditions or assumptions that lead to different behaviors. |
| Usage | Determine informally the time and space complexity of simple algorithms. |
| Familiarity | State the formal definition of big O. |
| Familiarity | List and contrast standard complexity classes. |
| Assessment | Perform empirical studies to validate hypotheses about runtime stemming from mathematical analysis Run algorithms on input of various sizes and compare performance. |
| Familiarity | Give examples that illustrate time-space trade-offs of algorithms. |
| Usage | Use big O notation formally to give asymptotic upper bounds on time and space complexity of algorithms. |
| Usage | Use big O notation formally to give expected case bounds on time complexity of algorithms. |
| Familiarity | Explain the use of big omega, big theta, and little o notation to describe the amount of work done by an algorithm. |
| Usage | Use recurrence relations to determine the time complexity of recursively defined algorithms. |
| Usage | Solve elementary recurrence relations, eg, using some form of a Master Theorem. |

**Learning Methods**

Figure 1:12 Learning Outcomes Based ACM/IEEE

# Chapter 2

## 1. Requirements Analysis

Many studies and projects which have been conducted worldwide on data integration were used in different sectors, and in many fields such as learning, education and health. In the following section we will study and compare these studies.

### 1.1. **Literature Review**

| **Project Name** | **Description** |
|---|---|
|  | **OpenScout:**<br>A European project co-funded by EU *e*Content*plus* Programme.<br>The project targeted the area of educational content and has18 partners in 13 countries, who are organized in 8 groups.<br>Duration: Sep. 2009 – Aug. 2012<br><br>**Open Scout develops tools & services for:**<br>- Skill-based scouting of<br>- Open content for<br>- Management education and training.<br><br>**Provides education services in the internet that enable users to** easily find, access, use and exchange open content for management education and training.<br><br>**OpenScout can be used by :**<br>- Learners, directly<br>- By training and education institutions that search for learning content to |

| | |
|---|---|
| | be integrated into their learning offerings.<br><br>**OpenScout envisaged wide and large scaled education service will consist of the following components:**<br>**Federated Content Base**<br>- Connect leading management content repositories.<br>- Extend with metadata to improve retrieval.<br><br>**Web Services:**<br>- Skill-based search & retrieval services.<br>- Integration into Learning Management Systems (LCMS) and social network platforms.<br><br>**Tool Library.**<br>Re-authoring, improvement, annotation of content<br>**Open Content Community**<br>Enlarge group of learners & providers adopting.<br>Services & standards for open content.<br>contribute metadata (such as skill-data,<br>User evaluations, ratings etc.)<br>Support design and evaluation of OpenScout's prototypes.<br>**Objectives**<br>**User-friendly access to open management content**<br>To exploit the full potential of open content for management education and training, OpenScout aims at providing an educational internet service with the following main objectives:<br>Enable user-friendly search and access to open management content.<br>Identify and connect existing repositories containing open management content.<br>Enable easy re-use and adaptation of open content to the need and context of learners and providers.<br>Enlarge the group of learners and providers that utilize open content for management education. |

| | |
|---|---|
|  | The use of outcome and competency frameworks is a growing part of health professions education and regulation in many countries.<br><br>The MedBiquitous Competency Framework, is a technical standard for representing competency frameworks in XML<br><br>**The standard allows medical schools and other health professions schools**<br>connect their curriculum,<br>learning resources,<br>Assessment data back to a common set of competencies, ultimately enabling competency-based views of the curriculum and of learner performance.<br><br>**To implement competency-based education and assessment, competency frameworks must be :**<br>Connected to curricula.<br>Learning resources.<br>Assessment data.<br><br>**Once competencies are expressed in a common format, they can be used as the backbone of education and performance management systems and offer a number of benefits.**<br><br>Learners and educators can search for learning resources addressing a particular competency.<br>Educators can determine where specific competencies are addressed in a curriculum.<br>Boards and hospitals can track and manage competency data for the professional.<br>Administrators can map one competency framework to another. |

| | |
|---|---|
| PROLIX | **Prolix** is a 48 months research and development integrated project co-funded by the **European Commission**.<br><br>The **Sixth Framework programme**, Priority 2 **Information Society Technologies**" started on December 1st, 2005.<br>**The objective of Prolix is to**<br>Align learning with business processes in order to enable organizations to faster improve the competencies of their employees according to continuous changes of business requirements.<br>To reach this goal, **PROLIX** develops open, integrated reference architecture for process-oriented learning and information exchange.<br><br>**PROLIX** supports a complete learning process life cycle:<br><br>The analysis of complex business situations.<br>The identification of individual and organizational learning goals.<br>The analysis of competencies and their matching with individual skills.<br>The definition of appropriate learning strategies and the simulation of competency-oriented processes.<br>The execution of improved learning processes.<br>The monitoring of learners' performance according to the goals defined. |
| ICOPER | ICOPER is the community of the 30 month eContentplus Best Practice Network ICOPER started in September 2008<br><br>The ICOPER networking capacity is grounded in key and leading experts in digital educational development and technology enhanced learning (TEL).<br><br>Driven by a consortium of 23 key players in Europe will provide access to a critical mass of more than 12,500 hours of integrated educational content. |

| | |
|---|---|
| | Based on this beneficial infrastructure the project will systematically analyze the specifications and standards available and in use, to draw conclusions on their validity in ICOPER Suitability Reports for Better Practice.<br><br>ICOPER's underlying educational framework will guide a consensus building approach to developing Best Practices, addressing issues such as:<br>Exchange of competency models and learning outcomes.<br>Collaboration around learning designs.<br>Integration of content via federated search and harvesting.<br>Reuse of instructional models and content in learning delivery environments.<br>Interoperability of item banks for assessment and evaluation.<br><br>ICOPER will provide mechanisms to ensure European-wide user involvement, cooperation, and adoption of standards within a large community to support all phases of standardization.<br><br>Overall, the confusion around the applicability (fit-for-purpose) of standards and specifications in technology enhanced learning results in a lack of adoption, which consequently has a profound negative impact on making digital content in Europe more **accessible**, **usable**, and **exploitable**.<br><br>The work in ICOPER is driven by an educational framework that is competency-driven and consists of **4 process stages** where best practices and use of specifications and standards are analyzed, and the result is integrated in the ICOPER reference model. |

Table 4 : Literature Review

The first step is to determine main components, features, target users and the system work flow according to the following:

1.2. **User Overview**

To define the users group we need to focus on the user's features:

1. Course learning outcomes;
2. Course objectives;
3. Specified courses (Course Category);
4. Instruction languages;
5. Location;
6. Prerequisites;
7. Course context;
8. Achieved learning outcomes;
9. Location of the holders;
10. Intended learning outcomes;
11. Intended learning outcomes;
12. Languages;
13. Related learning outcomes;
14. Assessment records result;
15. Assessment level and value;
16. Job description;
17. Learning level (Diploma, Bachelors, Master …);
18. Job Locations;
19. Experience years;
20. Language;
21. Skills and Knowledge.

### 1.3. **Users Classes**

In this section we need to determine the target user groups as the following:

1. Learners: persons who need learning opportunities.
2. Employers: persons who work or not, and who need to improve their skills and learning/or need jobs.
3. Learning Opportunities: places that provide the learning courses, such as universities, institutes or training centers.
4. Job Opportunities: places that provide jobs with specifications, such as Recruitment Companies or human resources departments.

Figure 2:1 Target Users

## 2. Business processes

### 2.1. **Processes Overview**

The core process of the system depends on three components; the first one is learning opportunity (courses), the second is job opportunity, and the third is what the individual has (Individual Profile). This component is called ‘personal achieved learning outcomes’, and the focus of this Thesis will be on the learning outcomes for every component to develop the matching engine.

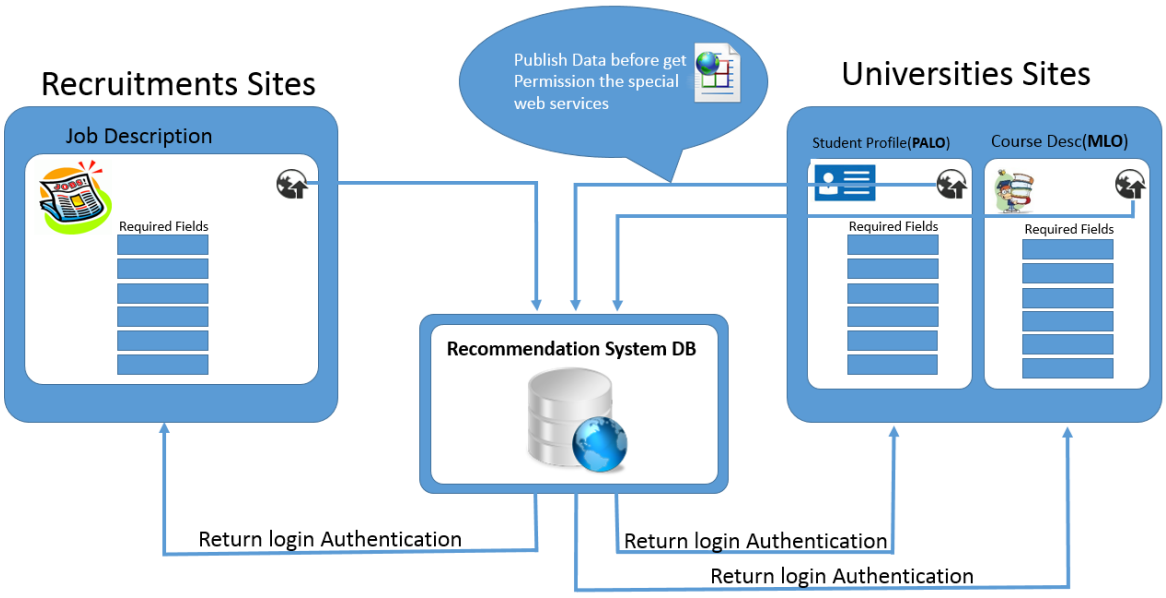



Figure 2:2 Matching Engine Workflow Scope

These components work together to assign or recommend a job or a course to the Individual, depending on this individual's achieved learning outcomes.

When students complete their university education or a training course, they give the university or institution the permission to publish their personal and education profile through many publishing techniques like web service, RSS or atom feed on the related local database system automatically, this type of publishing depends on PALO Standard that were previously explained in this research. The published data consists of personal information, which is the holder's name, Identifier (primary url that is used to access the profile) and description (a human readable description of the learner profile), achieved learning outcomes for each learner, and the intended learning outcomes attained by the learner. The system creates user profiles automatically when exporting the data from the source, and allows the learners to manage their achieved learning outcomes and other related items, such as working career path and learning path.

On the other side, when the universities or educational centers want to provide certain courses, they publish specifications of these courses through many publishing techniques

like web service, RSS or atom feed with many attributes like location, title of the course, description, instruction language assessments, prerequisites, teaching methods learning outcomes and objectives through MLO Standard that we explained at the beginning of this thesis. The system creates the automatic MLO provider to manage the information related to courses.

The last part is the job providers who declare the job qualifications with learning outcomes, skills, knowledge and other required attributes, automatically through web service or RSS. The Job Profile consists of the following items: job title, name of company or institution, the job application deadline, job description, required skills and knowledge.

The system can work with all learning outcomes types as input to give recommendation for job or course as output. This output displays the matching percentage for every person to help the employers choose the best candidates for the jobs and to arrange for the interviews, as the next step in the process. The diagram below displays the data flow from the sources to the local database and the relations between these components to accomplish the recommended functions.

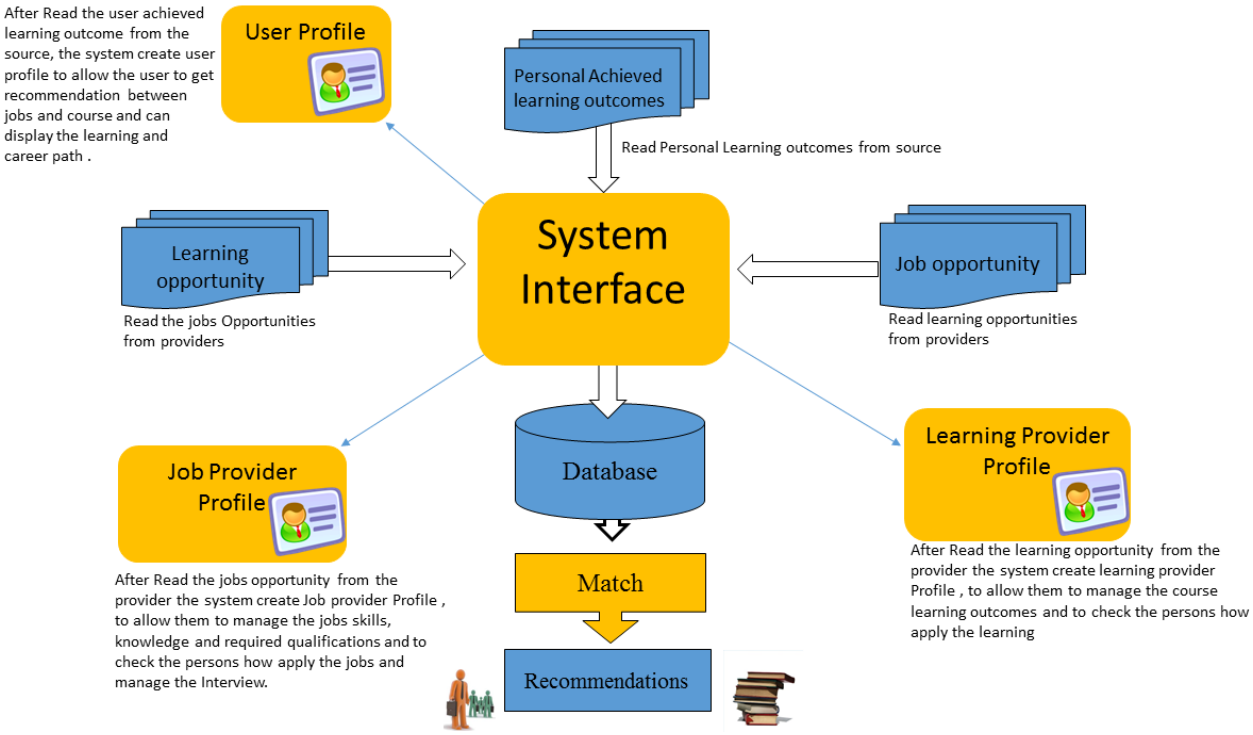



Figure 2:3 System Data Flow Diagram

The system allows learning opportunity providers to create profiles to manage the learning opportunities that they need to publish through the system. This feature can allow adding new opportunities from the provider directly, and adding the learning outcomes, objectives and learning method as shown in Fig. 2.4

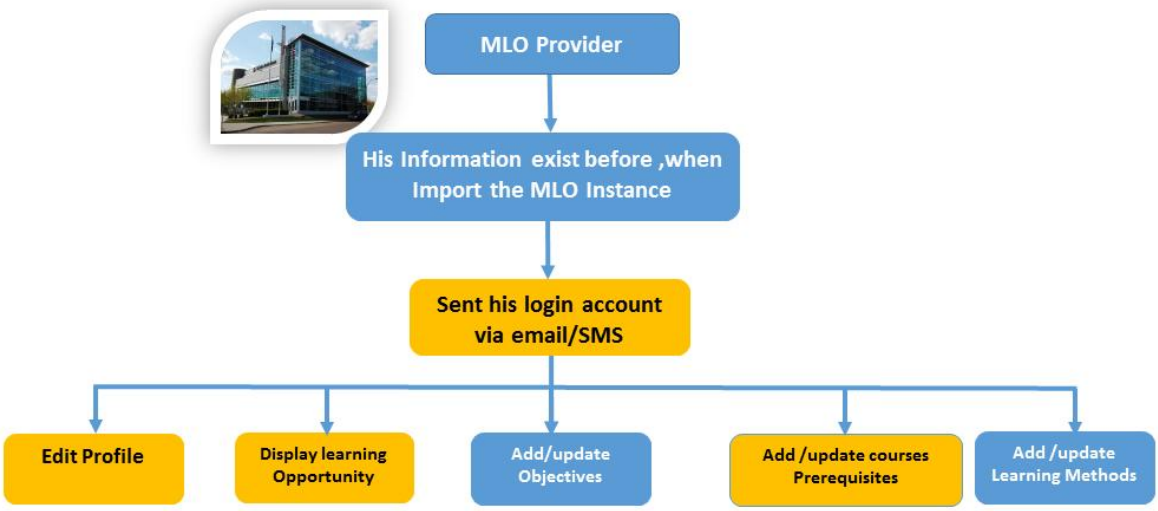



Figure 2:4 MLO Data Flow Diagram

Also, the job provider can add and manage the job opportunities directly from the source by creating profile; through this profile the provider can also manage the interview steps as shown in Fig. 3.5

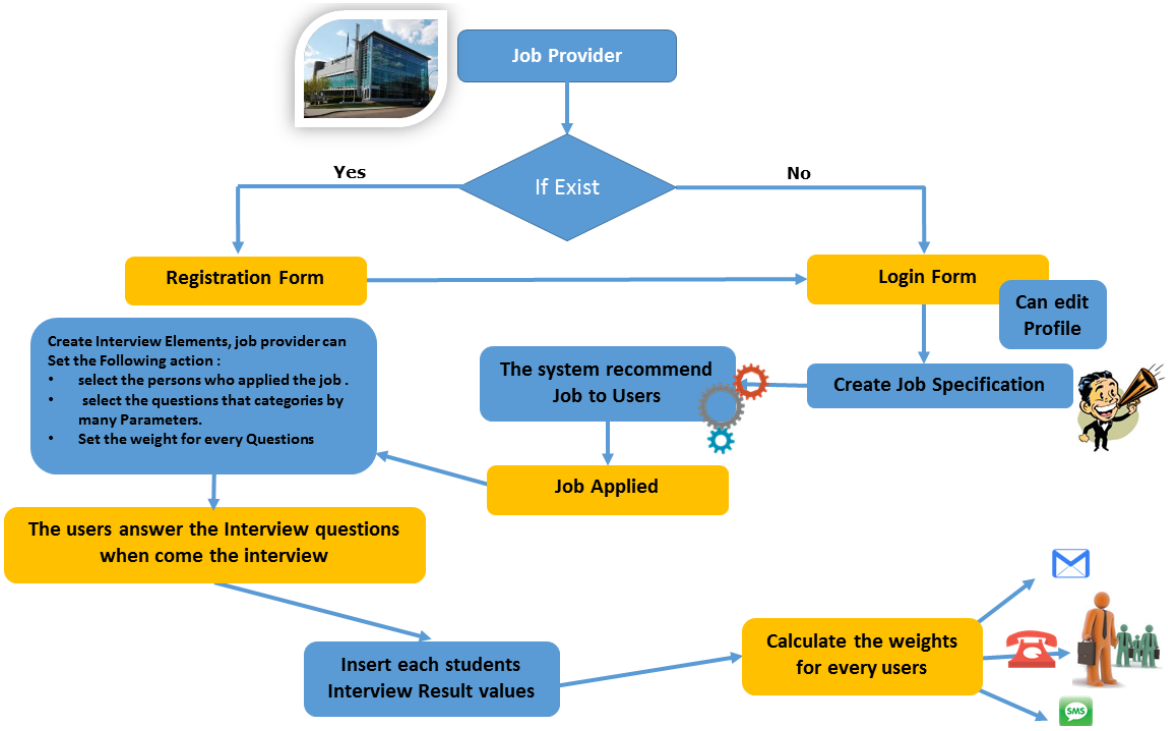



Figure 2:5 Job Dataflow Diagram

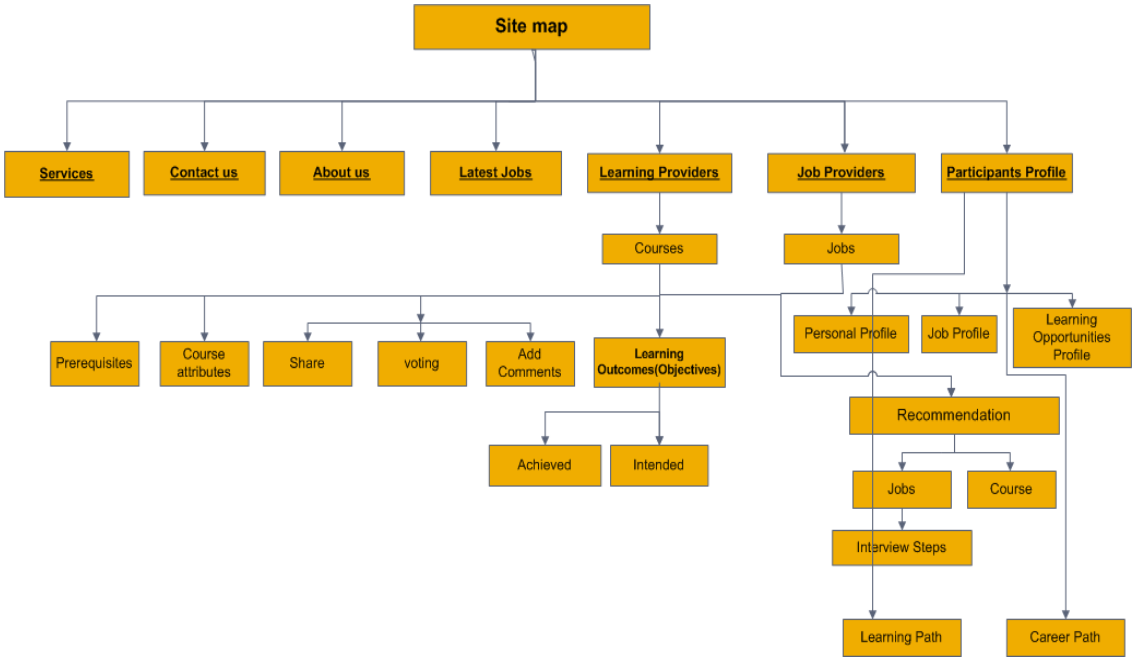



Figure 2:6 Main Parts in UI

## 2.2. **Job Opportunity**

After the Learners complete the learning opportunity by achieving the learning outcome of this opportunity, they begin to search for a job to meet their needs and achieve their ambitions; they take a lot of time to search at magazines newspapers and recruitment websites.

Sometimes they publish their CV through a variety of social networking site such as Facebook, Twitter and LinkedIn, maybe so lucky to get a job and maybe they take a lot of time and effort to search.

The new system keeps scientific qualifications of individuals up to date during the transmission of this data from universities and education centers directly to the system, without any human intervention, therefore data transmitted to the system is 100% identical to the data obtained from universities and education centers.

Once the recruitment company has the Profile in the system, they can declare new job vacancies which determine the required skills and knowledge, and can perform many

other tasks through the system, such as adding new job vacancy, getting information about applicants for any vacancy, managing interviews, setting interview questions and assigning weight for each one, preparing the interview and setting the interview results and allowing the company to select one more suitable person for specific jobs, as shown in Figure 2.7 below.

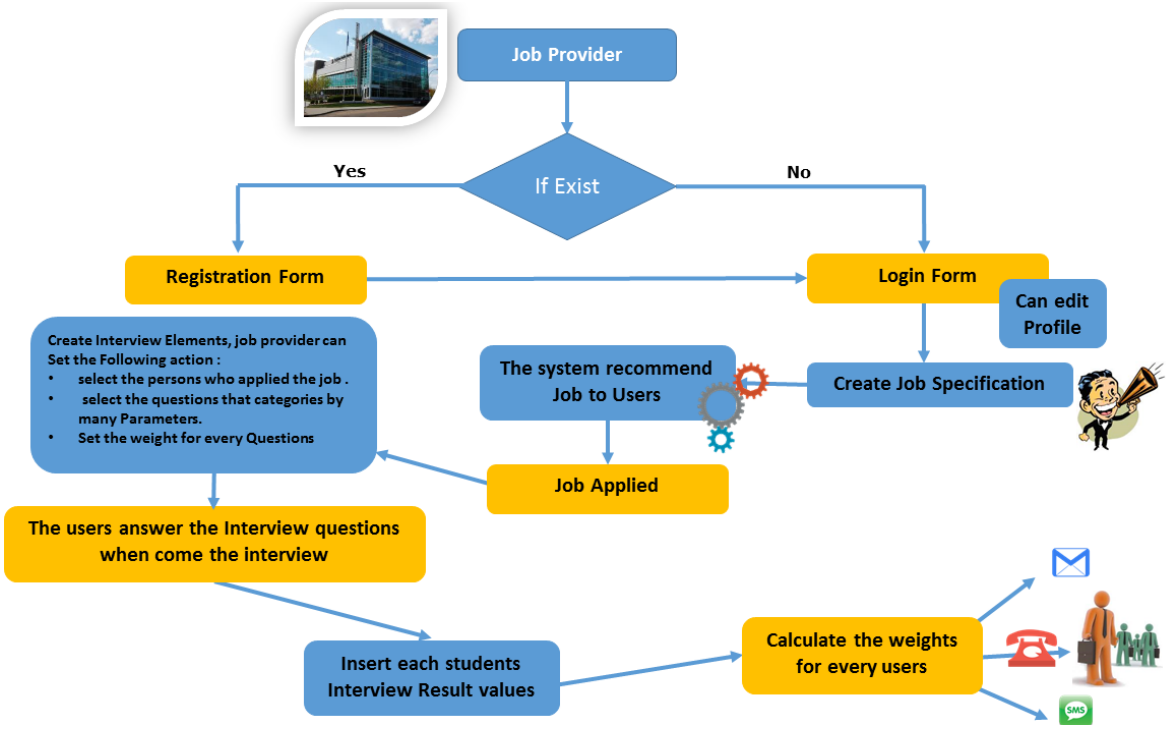



Figure 2:7 Interview Dataflow

## 2.3. **Interview Managements**

Job interviews are conducted by companies and organizations in order to select the best qualified human resources for a job. Interviews are formalized and performed by a specialists in order to assess a candidate for a job interview.

The aim of the interview is to assess the suitability of candidates for a specific job. Most organizations use the interview at some stage in their selection process.

There are different kinds of interviews through which qualified candidates are selected: telephone interviews, exam interviews and personal interviews. The last type is the most popular type of interview.

Interviews are useful for assessing the personal characteristics of a prospect candidates, such as practical intelligence, interpersonal skills and communication skills. The interview can be used for answering applicants' questions, selling the organization and negotiating terms and conditions.

When the system provides recommendations to jobseekers and they apply to the recommended job, the job provider administrator will be able to find the persons who applied to the job automatically, and can perform the following steps:

1. Create interview process;
2. Select the persons who accept this job;
3. Set interview questions and assign weights to each question. Questions can be divided into categories; each group of questions under a main category.
4. Conduct personal interviews and calculate scores for the answered question and enter them into the system.
5. The System calculates the interview results and displays the matching percentage of the job for each person, in order to enable the administrator to make the right decision.
6. The interview questions can be divided into many categories, each one has a set of sub-question. These questions have weight where the person can know when to answer the question during the interview. Each interview has one or more categories.
7. Adding the matching score to the question answering score as total to give recommendation on who is the suitable one for this interview.

8. The categories which are used to classify the questions are:

| Category Name(Parameter Name) | Description in Arabic |
|---|---|
| Motives | الدوافع |
| Marks | العلامات والتحصيلات |
| Experience and Responsibilities | الخبرات والمسؤوليات |
| Goals | الاهداف |
| Interest | الاهتمامات |
| About The Company | عن الشركة او المؤسسة |
| Leadership | القيادة |
| Money /Salary | المال والراتب |
| Flexibility | المرونة |
| Organization | التنظيم |
| Your preferences | خيراتك المفضلة |
| The ability to persuasion | القدرة على الاقناع |
| Qualifications | المؤهلات |
| Evolution | التطور |
| Travel \ workplace | السفر او مكان العمل |
| The ability to Sell | القدرة على البيع |
| Expulsion or termination | الطرد او انتهاء الخدمة |
| Evaluation | التقييم |
| Stimulus | التحفيز |
| Self-Assessment | التقييم الذاتي |
| Work under pressure | ضغط بيئة العمل |
| Perspectives of others | وجهات نظر الاخرين |
| Decision Making | اتخاذ القرارات |
| Languages | اللغات |

Table 5 : Interview Parameters

| نموذج تقييم مقابلات | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Q1 | هل لديك رخصة قيادة سارية المفعـــــول ؟؟؟ | | | | | | | | | | | | | | | | | |
| Q2 | هل تفضل العمل المكتبي او الميداني؟؟ | | | | | | | | | | | | | | | | | |
| م | الاسم | المظهر الخارجي | اللباقة | القدرة على التواصل و التعبير عن الافكار | الثقة بالنفس و قوة الشخصية | ENGLISH | COMPUTER | الاستعداد للعمل و العمل ساعات اضافية | الاقامة بالمؤسسة و الوظيفة | التنقلات و الدورات و الخبرات السابقة | المجموع | الراتب المتوقع | Q1 | Q2 | Q3 | Q4 | الترشيحية | |
| | الحد الاقصى للعلامات | 10 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 50 | الحد الادنى | تعدلا | تعدلا | تعدلا | مكتبي/ ميداني | قبول | اعتذار |
| 1 | | | | | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | | | | | | |

Figure 2:8 Evaluation Interview Form

# 3. Requirements Specification

## 3.1. **Functional Requirements**

The system can perform several operations by focusing on the functions and tasks of each one of the main components:

- **Learners**: individuals who need learning or jobs, this component can perform the following tasks:
    1. Login;
    2. Remember password;
    3. Check learning details;
    4. Edit profile;
    5. Explore achieved learning outcomes;
    6. Explore intended learning outcomes;
    7. Add comments to the course;
    8. Make voting;
    9. Share on social network Sites;
    10. Search learning outcomes;
    11. Apply course opportunity;
    12. Apply job opportunity.

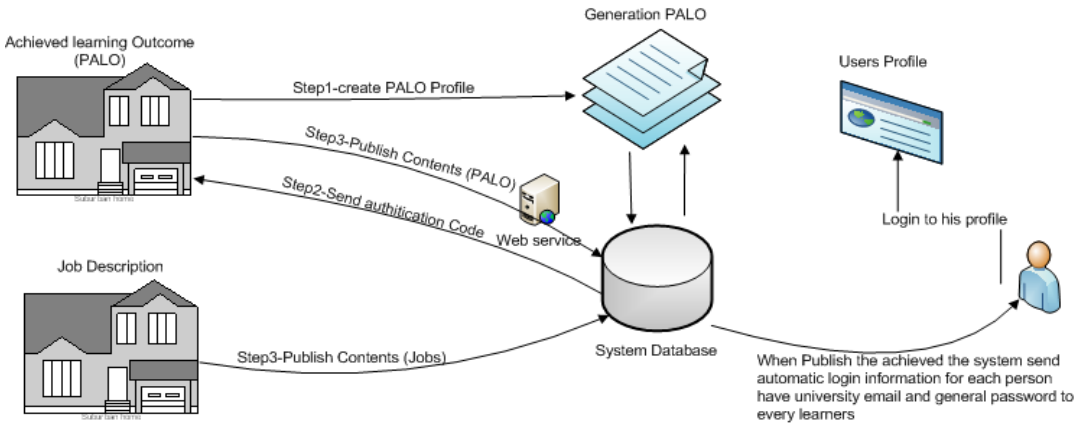



Figure 2:9 Personal Achieved Learning Outcomes Dataflow

- **Learning Providers**: the learning provider is the entity which provides the learning (course) like universities and institutes. This component can perform the following tasks:
    1. Login to the system;
    2. Manage provider profile;
    3. Manage learning attributes;
    4. Manage learning prerequisites;
    5. Manage learning outcomes for each instance;
    6. Manage applied users to specific course.

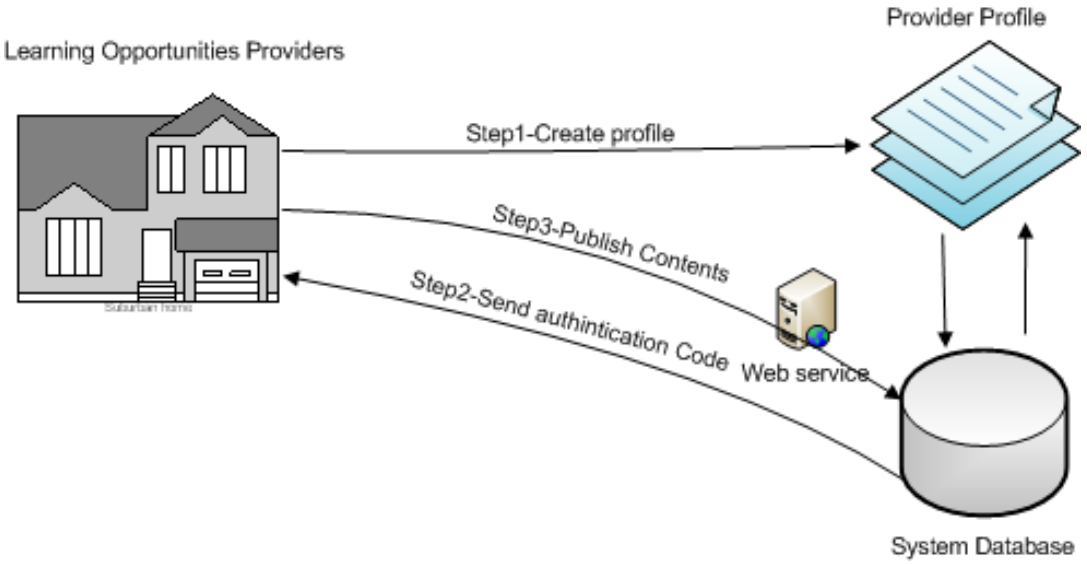



Figure 2:10 Learning Opportunity Dataflow

- **Job Providers**: Job providers are those who provide the job; such as human resource departments in companies and recruitment companies. They can perform the following tasks :
    1. Login to the system;
    2. Create job requirements (skill and knowledge);
    3. Create interview questions;

4. Select the persons who should be added to the interview process;
5. Set weight for each questions in the interview;
6. Conduct face to face interviews;
7. Set the interview results for each one;
8. Make the decision to select the most suitable candidate.

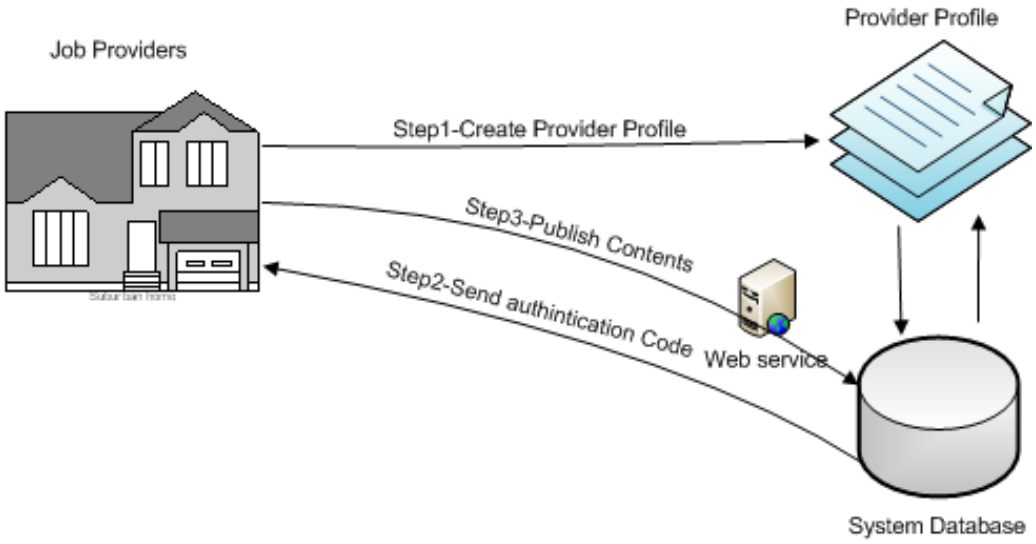



Figure 2:11 Job Opportunity Dataflow

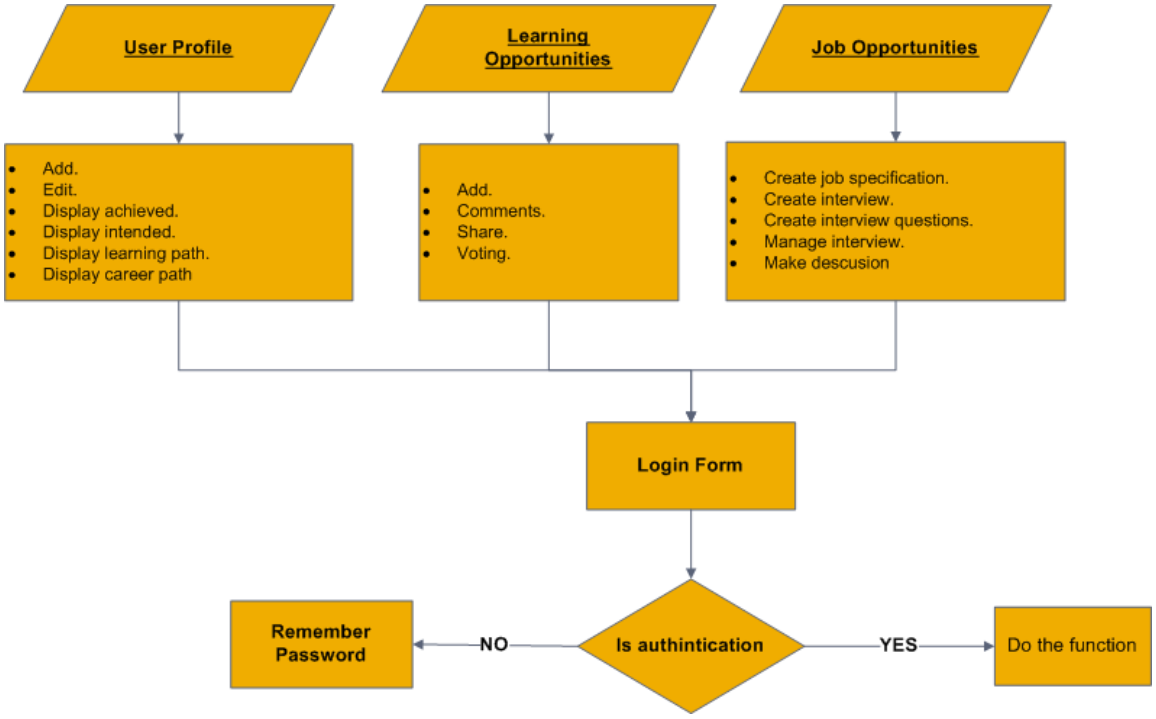



Figure 2:12 Required Authentication

### 3.2. **Non-functional Requirements**

1. Reduce the time and efforts required to get data;
2. Promote transparency within and between higher education systems;
3. Help graduate students to obtain employments appropriate to their field of specialty;
4. Provide accurate and up-to-date information on an individual's qualifications;
5. Provide training programs for graduate students to qualify for jobs;
6. Commodity information on courses and other learning opportunities, leveling the playing field and enabling new entrants to the market, e.g. Web 2.0 services.
7. Make the data it collected from universities and institutions accessible and centralized for employers;
8. Aid mobility and access to further studies and employment opportunities abroad;
9. Enable better consistency of information about learning opportunities across multiple services.
10. Provide fair and informed information regarding qualifications;
11. Support aggregator functionality such as rich browsing and targeted searches;
12. Link universities and institutions with each other as if they are a local network;
13. Facilitate academic and professional recognition and thus increase the transparency of qualifications;
14. Support the emergence of new business cases;
15. Take advantage of this data from other parties, such as libraries and centers of research and development;

16. Reduce the follow-up and supervision of data;

17. Secure;

18. Provide user validation inputs.

# 4. Object Modeling

## 4.1. **Objects**

An object represents a person, place, event, or transaction that is significant to the information system. The term “Object” usually refers to a particular instance; system analysts sometimes use the term to refer to a class of objects to consider how the UML describes the jobs.

The UML represents an object as a rectangle within the object name at the top, followed by the object’s attributes and methods.

We can create the Instances of the Jobs Description Object like:

| Jobs_Description | |
|---|---|
| **PK** | **JobID** |
| | JobIdentifier<br>JobTitle<br>CompanyName<br>Location<br>JobType<br>JobDescription<br>JobSkills<br>JobExperiance |

Figure 2:13 Object Attributes and Methods

C# is an instance from the Class job Description, and it has many attributes like working with framework 4.5 or higher, Job title: ASP.Net Products Developer, Job End Date and other attributes. Sometime database tables are used as classes and to create Objects from this class, the fields of the each table like the object attributes.

### 4.2. **Attributes**

Attributes are similar to adjectives that describe the characteristics of an object. For example: we have the object **course Info,** this object has many attributes like **course Name**, **course description**, **start date**, **location**, **cost**, **prerequisites**, **Objectives** and others.

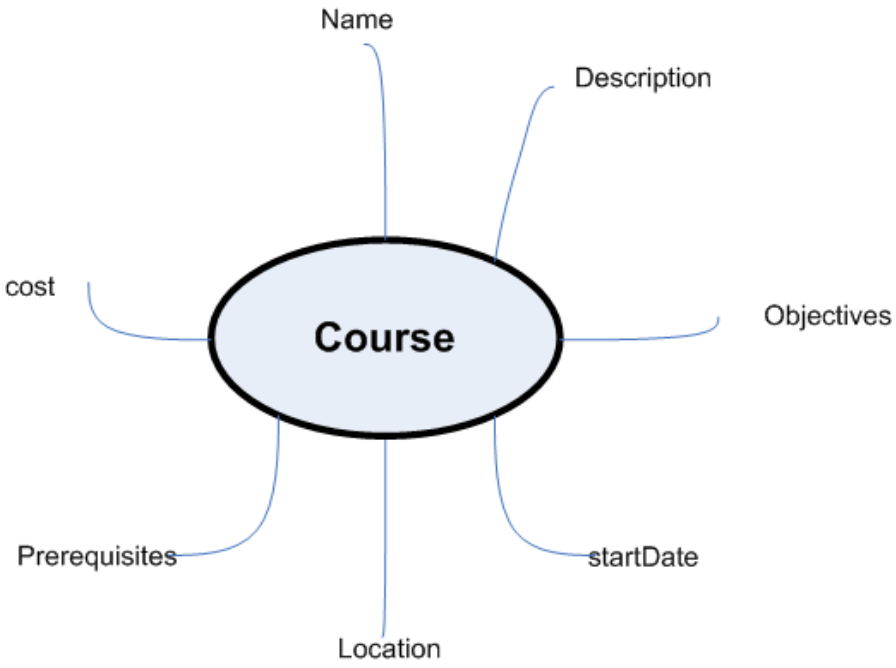



Figure 2:14 Course Attributes

### 4.3. **Methods**

A method defines specific tasks that an object can perform. Just as objects are similar to nouns and attributes are similar to adjectives, methods resemble verbs that describe what and how an object does something.

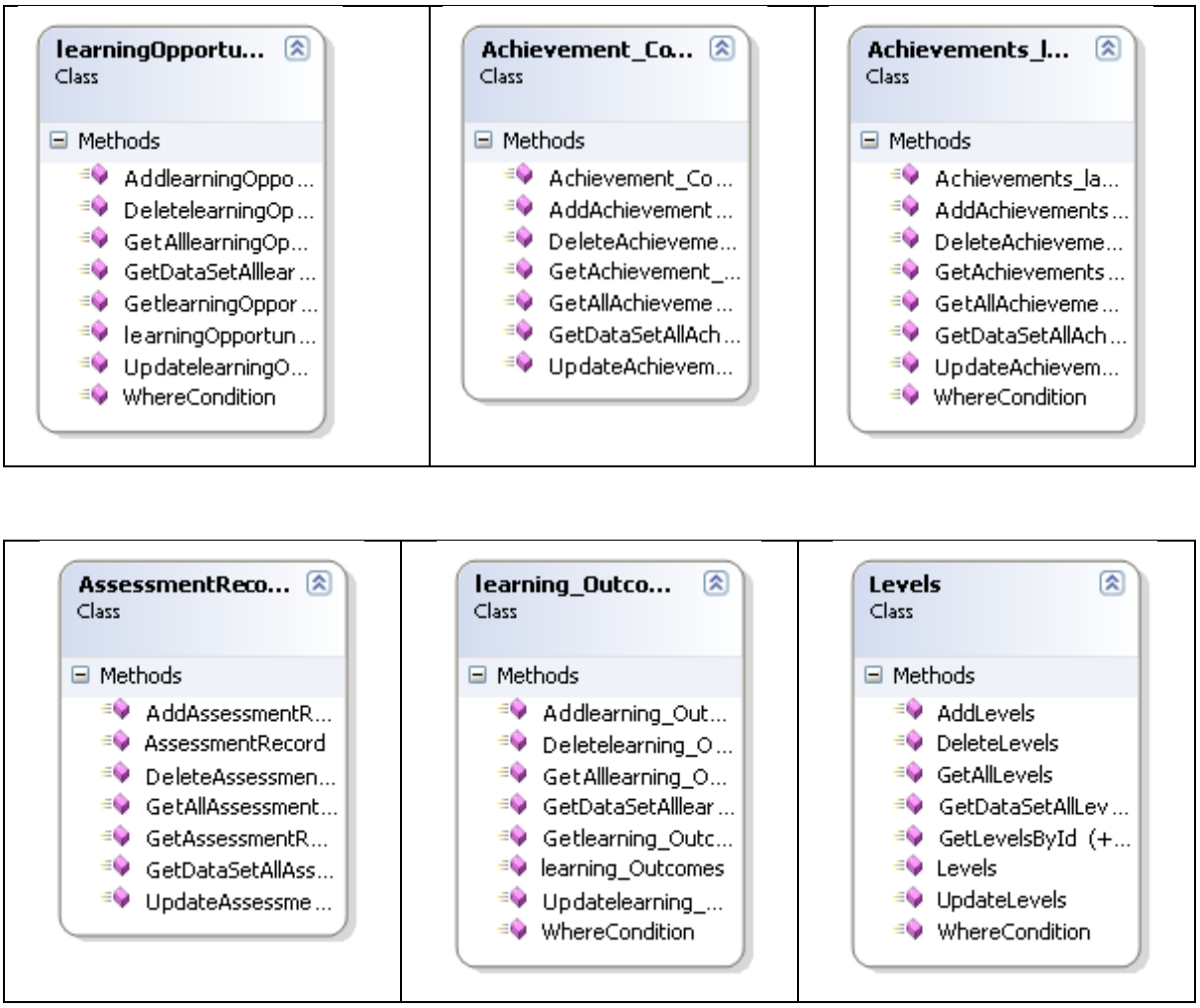


Figure 2:15 Methods

### 4.4. **Classes**

An object belongs to a group or category called a class. All objects within a class share common attributes and methods, so a class is like a blueprint, or template for all the objects within the class.

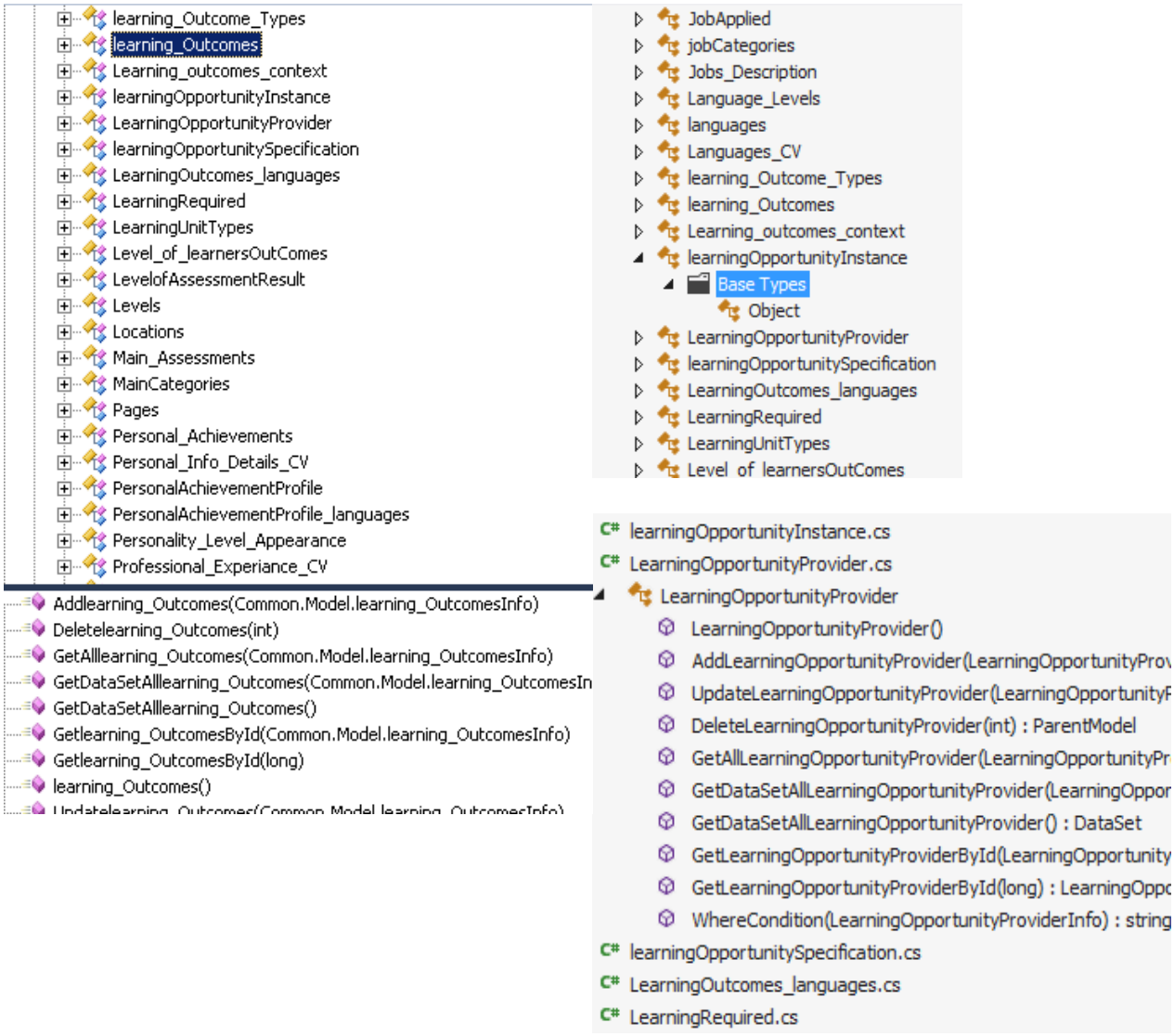



Figure 2:16 Classes

# 5. Object Relationship Diagram

## 5.1. **MLO Elements Relations**

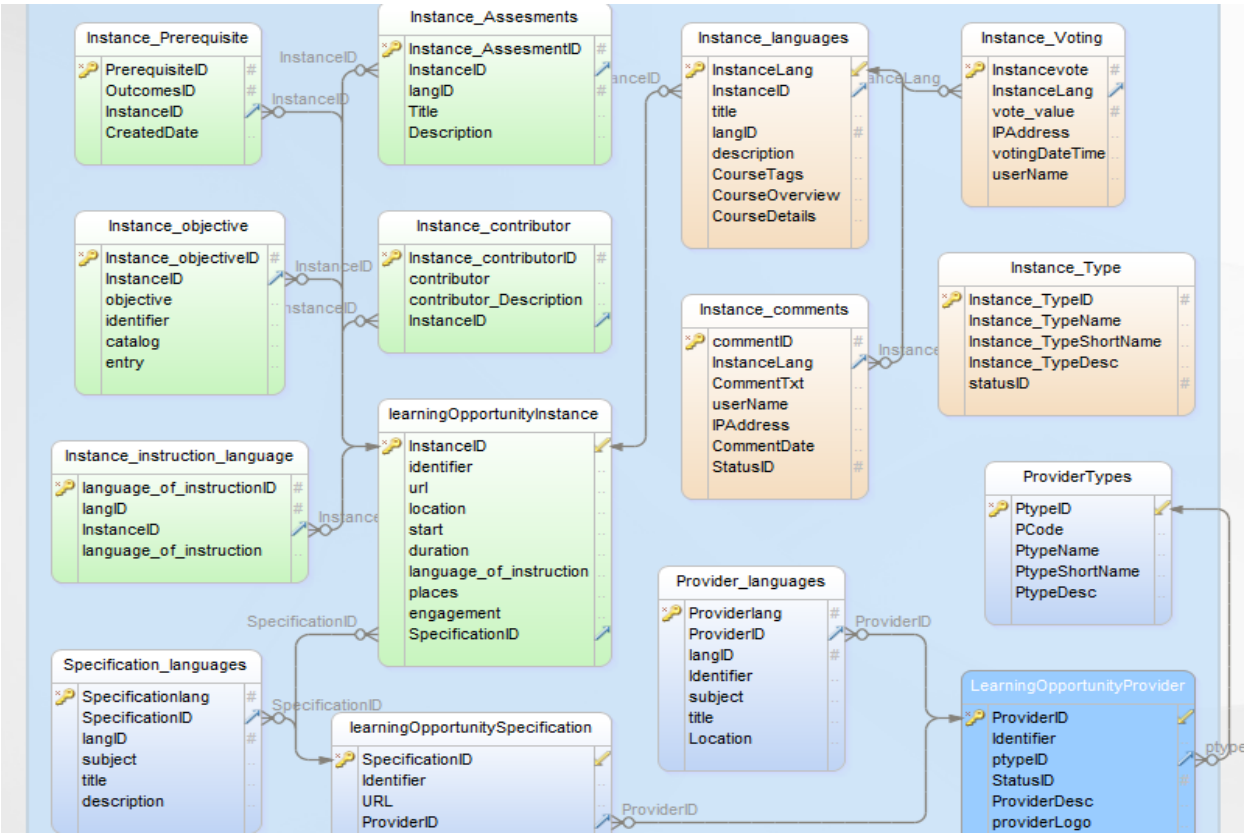



Figure 2:17 Objects Relations - MLO

## 5.2. **PALO Elements:**

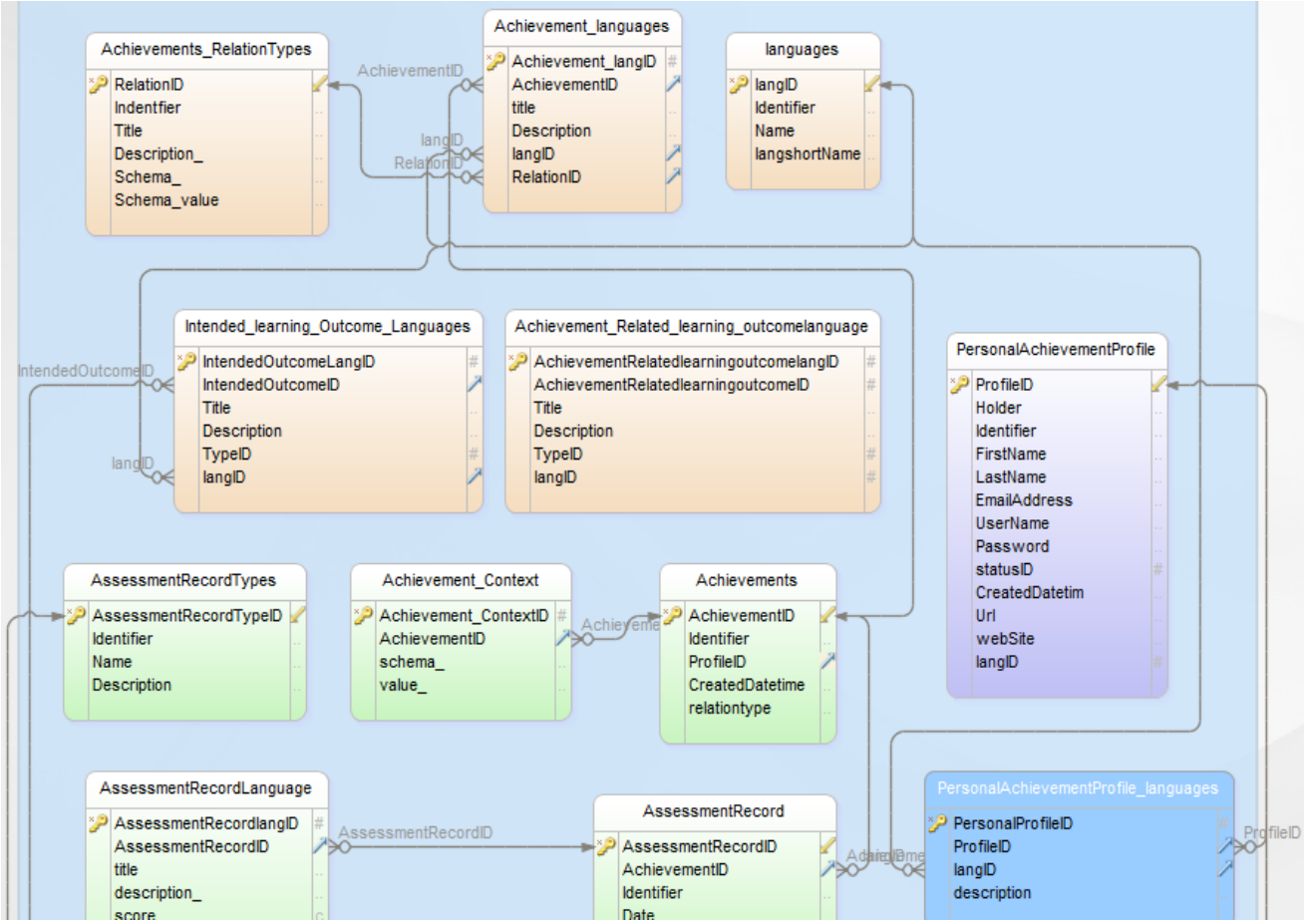



Figure 2:18 Object Relation - PALO

## 5.3. **Job Elements:**

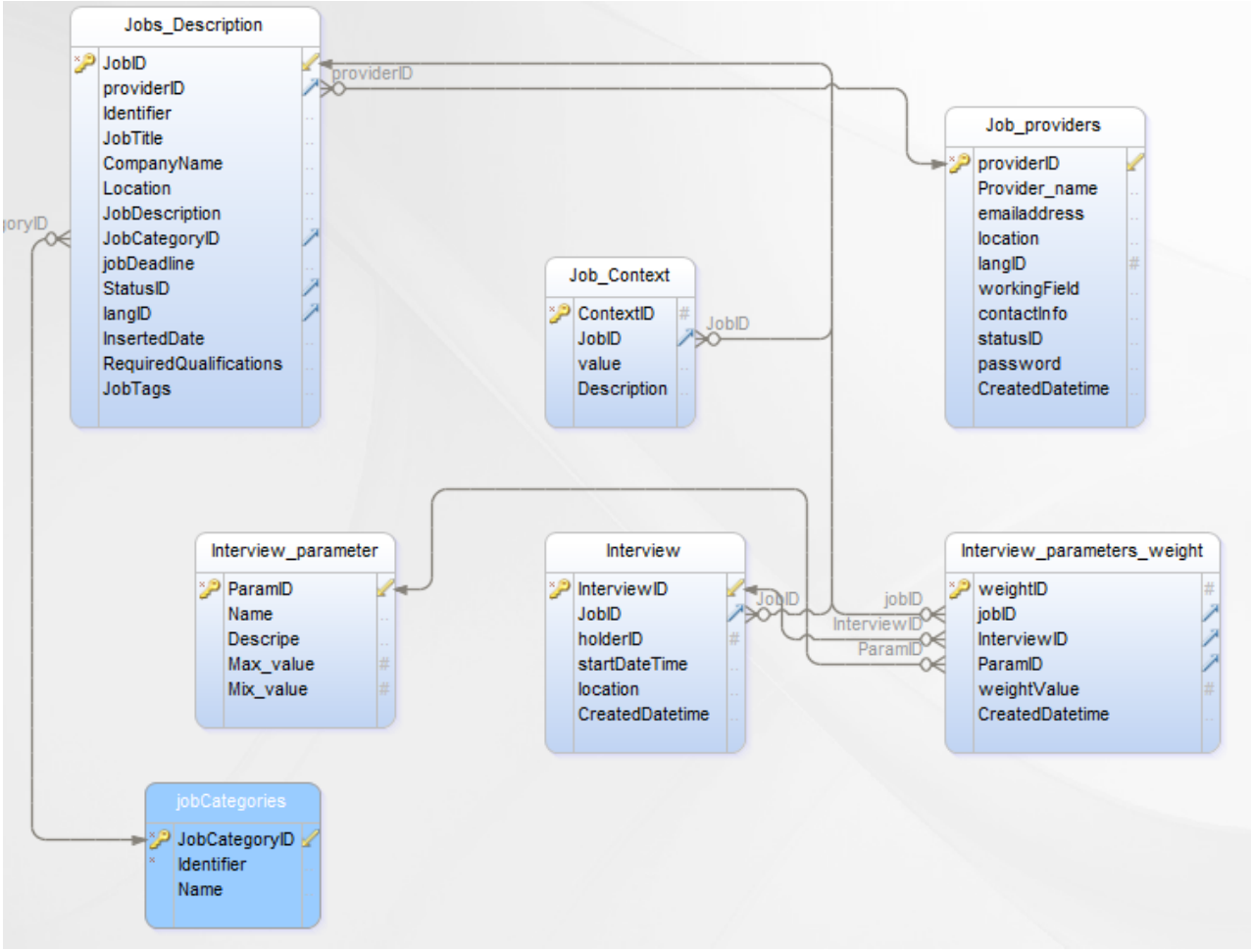



Figure 2:19 Job Tables and Relations

# 6. Matching Engine

## 6.1. **Introduction**

In order to assign jobs opportunities or learning opportunities to a particular participant, we need to create a recommendation system that depends on the main components (JOB/MLO/PALO) attributes. This recommendation system can provide the participant with suitable recommendations based on the content related to this participant (Content-Based). In content-based recommendations the system tries to recommend an item that matches the User Profile. The Profile is based on the archived learning outcomes that come from PALO.

A content-based recommender system matches the profile of the item to the user profile to decide on its relevancy to the user. The recommender system uses additional data about the context of item consumption.

To accomplish this recommendation, we need to analyze the main component attributes or parameters to create the automatic search engine, and this engine is based on the string matching results for this parameters**.**

| **MLO** | **PALO** | **Job Description** |
|---|---|---|
| Courses Learning Outcomes. | Achieved learning outcomes. | Job Description |
| Course Objectives. | Location of the Holders. | Learning level (Diploma, Bachelors, Master …). |
| assessment | Intended learning outcomes. | Job Locations |
| Instruction Languages | Languages | Experience yeas. |
| Location | Related learning outcomes. | Language |
| Prerequisites | Assessment Records Result. | Skills and Knowledge |
| Context | Assessment Level and Value | |

| | | |
|---|---|---|

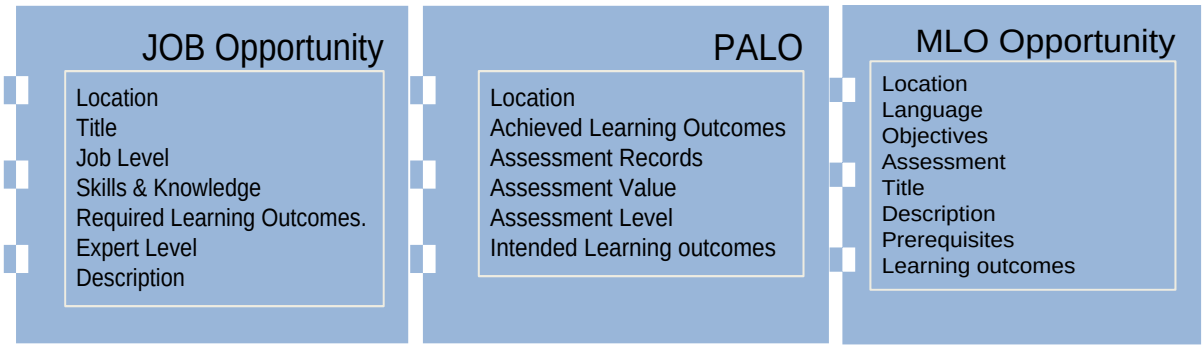



Table 2.3 Main Components Parameters

### 6.2. **How a Search Engine Works**

Search engine is the popular term for an information retrieval (IR) system. While researchers and developers take a broader view of IR systems, consumers think of them more in terms of what they want the systems to do. Search the Web, or an intranet, or a database.

Actually, consumers would really prefer a finding engine, rather than a search engine.

### 6.3. **Structured Data and Unstructured Data**

Items that can be recommended to the user are often stored in a database table. Table 10.1 shows a simple database with records. The column names (Job ID, Job Title and location) are properties of Job Description. These properties are also called attributes, fields, or "variables" in different publications. Each record contains a value for each attribute. A unique identifier, ID in Table, allows items with the same Job Title to be distinguished and serves as a key to retrieve the other attributes of the record.

| Job ID | Job Title | Location |
|---|---|---|
| 1 | ASP.NET 4.0 Developer | Ramallah |
| 2 | ASP.NET Senior Programmer | Jordan |

| | | |
|---|---|---|
| 3 | Accountant | Riyadh- Saudi Arabia |
| 4 | Asp.net | Ramallah |
| 6 | Php Developer | Jordan |

Table 2.4 Structured Data format

The Table above could be used to create a website that lists and recommends Jobs. This is an example of **structured data** in which there is a small number of attributes, each item is described by the same set of attributes, and there is a known set of values that the attributes may have. Many machine learning algorithms may be used to learn a user profile.

**Unstructured data** may occur in news articles, the entire article can be treated as a large unrestricted text field, such as:

| |
|---|
| (*Google debuted its latest Nexus devices—the Nexus 6, a huge new phablet with a 6-inch screen, and the Nexus 9, an 8.9 inch tablet—and formally launched Android 5.0 Lollipop, the latest version of its Android operating software.)* |

Table 2.4 Unstructured Data format

Unrestricted texts such as news articles are examples of unstructured data. Unlike structured data, there are no attribute names with well-defined values. Furthermore, the full complexity of natural language may be present in the text field including polysemous words (the same word may have several meanings) and synonyms (different words may have the same meaning).

Many domains are best represented by semi-structured data in which there are some attributes with a set of restricted values and some free-text fields. A common approach to dealing with free text fields is to convert the free text to a structured representation

Many personalization systems that deal with unrestricted text use a technique to create a structured representation that originated with text search systems.

### 6.4. **Items Representation**

Search engines match queries against an index that they create. The index consists of the words in each document, plus pointers to their locations within the documents. This is called an inverted file. A search engine or IR system comprises four essential modules:

1. A document processor;
2. A query processor;
3. A search and matching function;
4. A ranking capability.

### 6.5. **Document Processor**

The document processor prepares, processes, and inputs the documents that users search against. Comparing a document with a query requires both to be represented in a compatible form. A scoring function is used to assign to each document in the collection a value reflecting its relevance with respect to the current query.

**Tokenization**: It is common practice to use terms as the features to describe documents. To do so, each document needs to be individually parsed. Tokens, which are strings separated by whitespace or punctuation, are extracted by the use of a lexical analyzer. This step, known as tokenization, also converts strings to lower-case.

**Deleting stop words**: This step helps save system resources by eliminating from further processing, as well as potential matching, those terms that have little value in finding useful documents in response to a customer's query. This step used matters much more than it does now when memory has become so much cheaper and systems so much faster, but since stop words may comprise up to 40 percent of text words in a document, it still has some significance. A stop word list typically consists of those word classes known

to convey little substantive meaning, such as articles (*a, the*), conjunctions (*and, but*), interjections (*oh, but*), prepositions (*in, over*), pronouns (*he, it*), and forms of the "to be" verb (*is, are*). To delete stop words, an algorithm compares index term candidates in the documents against a stop word list and eliminates certain terms from inclusion in the index for searching.

**Stemming**: After a document has been reduced to a series of tokens and deleting stop words, stemming can be applied. The goal of stemming is to create a term that reflects the common meaning behind words such as “compute,” “computation,” “computer” “computes” and “computers.” Doing so reduces the number of unique tokens seen in the collection and has been shown to improve performance. The most common stemming algorithm is the Porter stemmer. This stemmer uses a long list of rules, hand-crafted for the English language, which successively removes commonly occurring suffixes.

## 7. System Algorithm

### 7.1. **String Algorithm:**

Machine learning is a fascinating area of computer science in which we study the techniques and implementations of algorithms that allow computers to learn. Its implementations include, but are not limited to, automatic text classification, search engines, stock market analysis, financial threshold alerts, speech and handwriting recognition, medical diagnostics, data mining, and autonomous land vehicles.

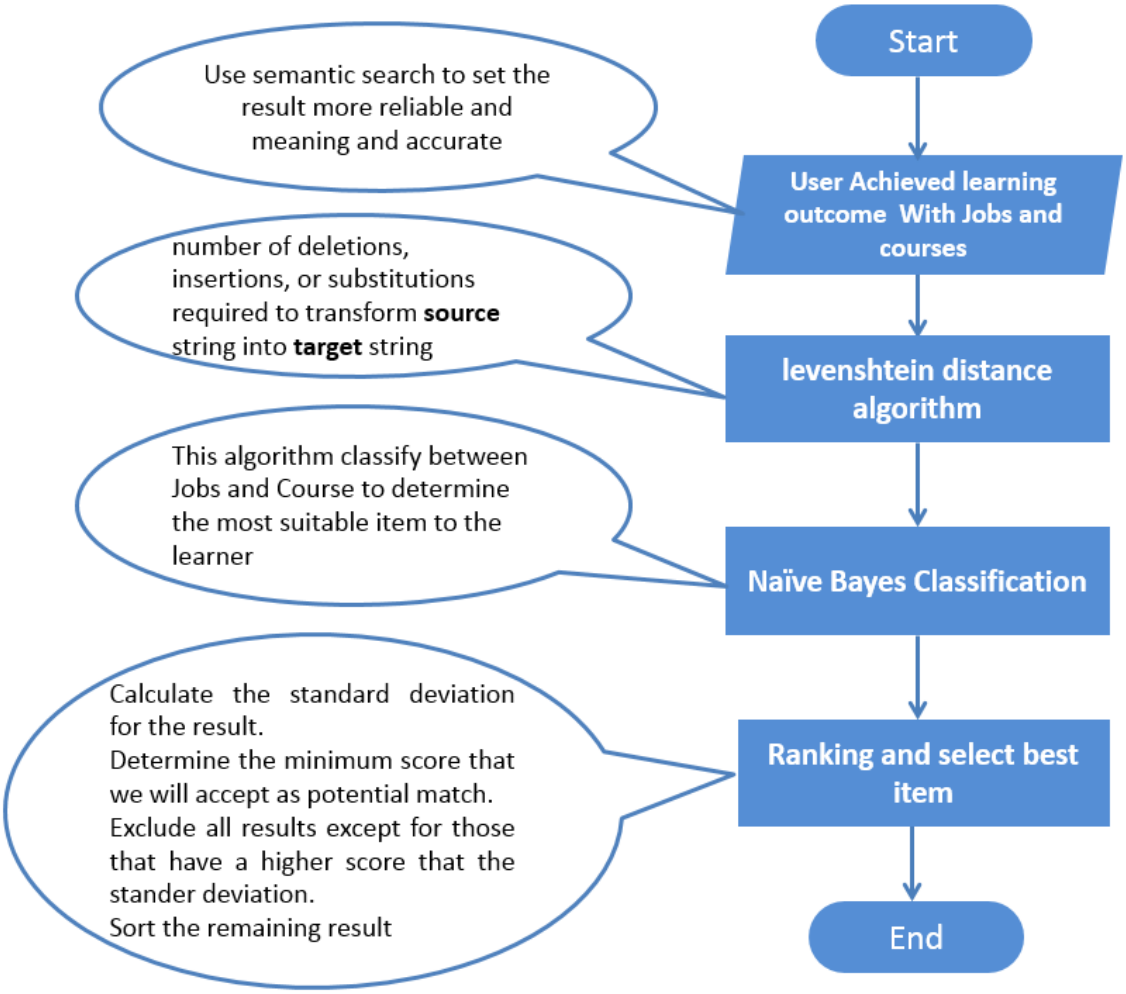



Figure 2:20 Required Matching Process

### 7.2. **Semantic Search**

Before start with the algorithm search and how they manipulate the text, we will start with and use semantic search to directly get what we need, its definition and its importance in this research.

**Semantic search** is a data searching technique in which a search query aims at not only finding keywords, but also at determining the **intent** and **contextual** meaning of the words a person is using for search. **Intent**, which comes from the user, explicitly states what he or she is looking for, and **context** could be understood as everything that surrounds a search and makes this go in either direction, what gives its meaning. Thus, by understanding and

connecting intention and context, search engines are able to understand the different queries, both what motivates and what is expected of them.

If we do Google search for **mint** keyword without specifying any description or other additional key words, the result displays for different meaning without understanding the meaning or the aim of the search and the result is not clear. One of the results means mint plant, mint leaves, mint newspapers or mint Linux, we see all these results to determine what we need specifically or what we search for, and the search engine returns different categories of search without specifying what we need exactly. See Google search results.

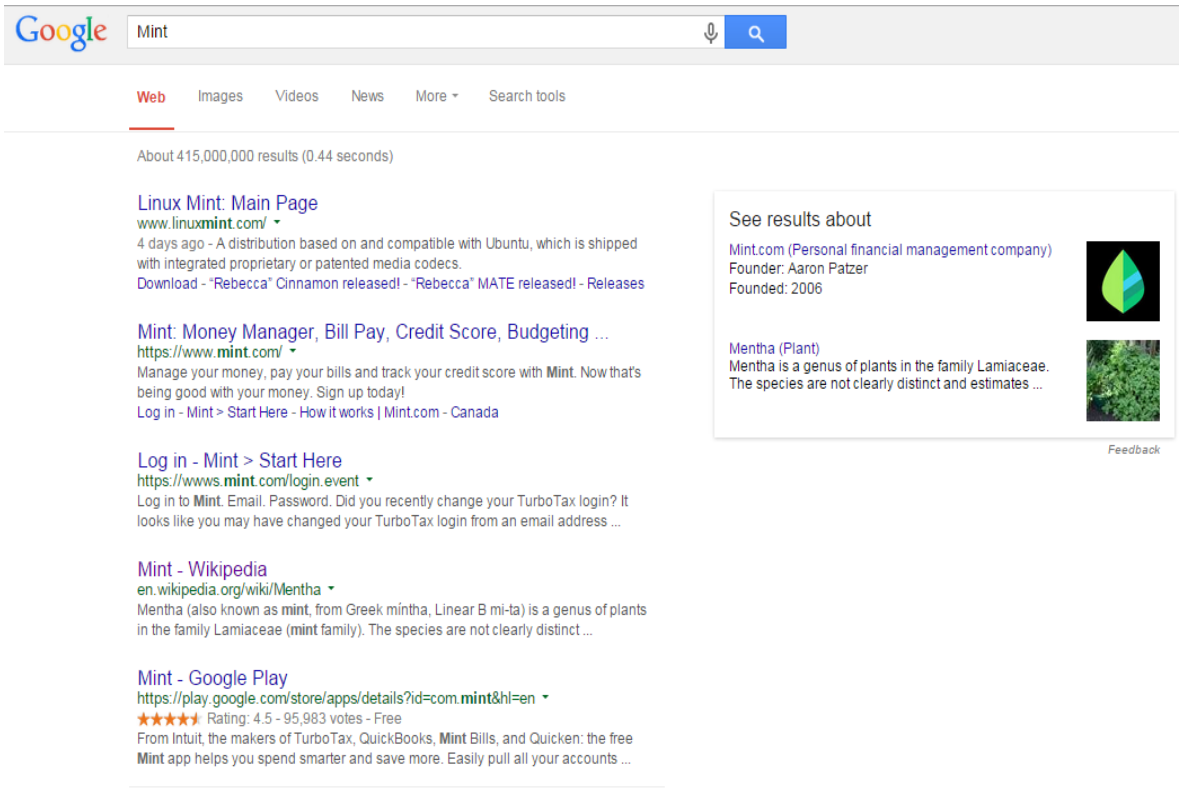



Figure 2:21 Google Search 'Mint' Keyword

In order to get more useful and clear results, we need to make the search depending on some features to get the aim result directly, and make the result more useful, clear and meaningful. Other keywords need to be added in order to classify the result, when adding a new search by typing in the search box ‘mint plant’, it is clear that Google search will

know perfectly the kind of results to show: those related to the plant, leaves. We find the following result:

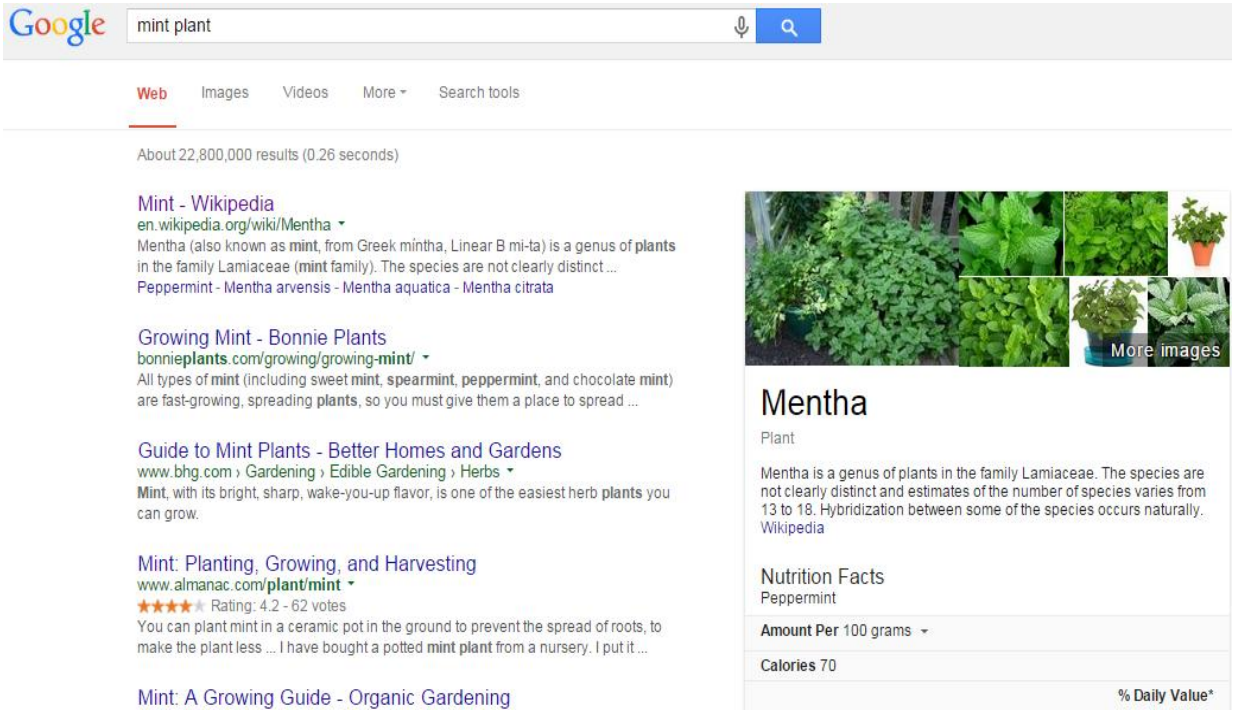



Figure 2:22 Google Search, More Specification

In the developed search engine, we need to apply the semantic search before starting the matching process in order to get clear and meaningfully results for Learners. Each learner has achieved different learning outcomes and have different characteristic or attributes, such as:

1. Gender;
2. Specification;
3. Location;
4. Languages;
5. Education Experience;
6. Learning level.

These attributes will be used for obtaining clearer and more accurate matching result.

### 7.3. **Levenshtein Distance Algorithm**

Levenshtein distance (LD) is a measure of the similarity between two strings, which we refer to as the source string (s) and the target string (t). The Levenshtein Distance algorithm is named after the Russian scientist Vladimir Levenshtein, who devised the algorithm in 1965. The Levenshtein edit distance is the number of insertions, deletions, or replacements of single characters that are required to convert one string to the other.

A distance between two “strings” can be computed by measuring the number of steps needed to turn one string into another:

distance(“palestine”, “palestin”) => 1 (deletion)

distance(“palestine”, “npalestine”) => 1 (insertion)

distance(“palestine”, “nalestine”) => 1 (substitution)

distance(“palestine”, “paletsine”) => 2

The Levenshtein distance algorithm has been used in:

1. Spell checking;
2. Speech recognition;
3. DNA analysis;
4. Plagiarism detection.

The Algorithm Steps are as follows:

| Step | Description |
|---|---|
| 1 | Set n to be the length of s.<br>Set m to be the length of t.<br>If n = 0, return m and exit.<br>If m = 0, return n and exit.<br>Construct a matrix containing 0..m rows and 0..n columns. |

| | |
|---|---|
| 2 | Initialize the first row to 0..n.<br>Initialize the first column to 0..m. |
| 3 | Examine each character of s (i from 1 to n). |
| 4 | Examine each character of t (j from 1 to m). |
| 5 | If s[i] equals t[j], the cost is 0.<br>If s[i] doesn't equal t[j], the cost is 1. |
| 6 | Set cell d[i,j] of the matrix equal to the minimum of:<br>a. The cell immediately above plus 1: d[i-1,j] + 1.<br>b. The cell immediately to the left plus 1: d[i,j-1] + 1.<br>c. The cell diagonally above and to the left plus the cost: d[i-1,j-1] + cost. |
| 7 | After the iteration steps (3, 4, 5, 6) are complete, the distance is found in cell d[n,m]. |

Table 6 : Levenshtein Distance Algorithm

Levenshtein distance computations:

Words: sql, sdql

Levenshtein distance: 1

Note: Only 1 edit is needed.

The'd' must be added at index 2.

Words: sqlserver, sql

Levenshtein distance: 5

Note: The final 5 letters must be removed.

Words: SQL2012, LSQ2012

Levenshtein distance: 3

Note: The first 3 letters must be changed

Drug names are commonly confused.

**C# code**

```
using System;
static class LevenshteinDistance
{
    public static int Compute(string s, string t)
    {
            int n = s.Length;
            int m = t.Length;
            int[,] d = new int[n + 1, m + 1];
            // Step 1
            if (n == 0)
            {
                return m;
            }
            if (m == 0)
            {
                return n;
            }
            // Step 2
            for (inti = 0; i<= n; d[i, 0] = i++)
            {
            }
            for (int j = 0; j <= m; d[0, j] = j++)
            {
            }
```

```
            // Step 3
            for (inti = 1; i<= n; i++)
            {
            //Step 4
               for (int j = 1; j <= m; j++)
               {
                        // Step 5
                        int cost = (t[j - 1] == s[i - 1]) ? 0 : 1;
                        // Step 6
                        d[i, j] = Math.Min(
                        Math.Min(d[i - 1, j] + 1, d[i, j - 1] + 1),
                            d[i - 1, j - 1] + cost);
               }
            }
            // Step 7
            return d[n, m];
   }
}
```

### 7.4. **Introduction for Text Categorization Applications**

Text can be represented in two separate ways. The first is as a bag of words, in which a document is represented as a set of words, together with their associated frequency in the document. Such representation is essentially independent of the sequence of words in the collection. The second method is to represent text directly as strings, in which each

document is a sequence of words. Most text classification methods use the bag-of-words representation because of its simplicity for classification purposes.

The goal of text categorization is the classification of documents into a fixed number of predefined categories. Text categorization is the task of automatically sorting a set of documents into categories from a predefined set, some of examples we are using text categorization include:

1. Web pages:
    1.1. Recommending;
    1.2. Yahoo-like classification.
2. Newsgroup/Blog Messages:
    2.1. Recommending;
    2.2. spam filtering;
    2.3. Sentiment analysis for marketing.
3. News articles:
    3.1. Personalized newspaper.
4. Email messages:
    4.1. Routing;
    4.2. Prioritizing;
    4.3. Folderizing;
    4.4. Spam filtering;
    4.5. Advertising on Gmail.

Automated text classification is attractive because it frees organizations from the need of manually organizing document bases, which can be too expensive, or simply not feasible given the time constraints of the application or the number of documents involved. The accuracy of modern text classification systems rivals that of trained human

professionals, thanks to a combination of information retrieval (IR) technology and machine learning (ML) technology.

Text classification is the task of classifying text documents to multiple classes, such as:

1. Is this mail a spam?
2. Is this Job/course suitable for x or y Learners?
3. Is this Job/Course likely to be relevant to user X?
4. Is this book possibly of interest to the user?

### 7.5. **Text Categorization Methods**

Text Categorization Methods are methods that sum evidence from many or all features (e.g. naïve Bayes, KNN, neural-net, SVM) tend to work better than ones that try to isolate just a few relevant features (decision-tree or rule induction).

- **Decision Trees**: Decision trees are designed with the use of a hierarchical division of the underlying data space with the use of different text features. The hierarchical division of the data space is designed in order to create class partitions, which are more skewed in terms of their class distribution. For a given text instance, we determine the partition that it is most likely to belong to, and use it for the purposes of classification.
- **Pattern (Rule)-based Classifiers**: In rule-based classifiers, we determine the word patterns which are most likely to be related to the different classes. We construct a set of rules, in which the left hand side corresponds to a word pattern, and the right-hand side corresponds to a class label. These rules are used for the purposes of classification.
- **SVM Classifiers**: SVM Classifiers attempt to partition the data space with the use of linear or non-linear delineations between the different classes. The key in such

classifiers is to determine the optimal boundaries between the different classes and use them for the purposes of classification.

- **Neural Network Classifiers**: Neural networks are used in a wide variety of domains for the purposes of classification. In the context of text data, the main difference for neural network classifiers is to adapt these classifiers with the use of word features. it is noted that neural network classifiers are related to SVM classifiers; indeed, they both are in the category of discriminative classifiers, which are in contrast with the generative classifiers
- **Bayesian Classifiers**: In Bayesian classifiers (also called generative classifiers), we attempt to build a probabilistic classifier based on modeling the underlying word features in different classes. The idea is then to classify text based on the posterior probability of the documents belonging to the different classes on the basis of the word presence in the documents.

The Naive Bayesian technique is the most popular and is used in contemporary spam filtration, document categorization, and help desk systems because of its accuracy, robustness, and ease of implementation. The Bayesian classifier is attributed to a British mathematician, Thomas Bayes who defined probability as:

***'The probability of any event is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the chance of the thing expected upon its happening'.***

Bayesian provides an efficient way to attain reason from uncertainty.

Example: Have you ever wondered what happens behind the scenes when you receive a spam message in your e-mail inbox and you click on the Mark as spam or Send to Junk mail button? The message gets passed to an indexer that tokenizes the contents, calculates

the probability, and adds it to a spam repository memorizing that the particular message had junk mail attributes. Also, for good messages, it learns from the environment and classifies the text as being not junk. When the next message comes in, the system automatically calculates the prior probability on the past experiences and can recommend if the new message falls under the category of ham or spam. This learning is based on the category of the word and its frequency of occurrence.

### 7.6. **Naive Bayes Variation**

There are several Naive Bayes Variations. Here we will discuss three of them:

1. The Multinomial Naive Bayes: used when the multiple occurrences of the words matter a lot in the classification problem. Such an example is when we try to perform Topic Classification.
2. The Binarized Multinomial Naive Bayes: The Binarized Multinomial Naive Bayes is used when the frequencies of the words don't play a key role in our classification. Such an example is Sentiment Analysis, where it does not really matter how many times someone mentions the word "bad" but rather only the fact that he does.
3. The Bernoulli Naive Bayes: Bernoulli Naive Bayes can be used when in our problem the absence of a particular word matters. For example Bernoulli is commonly used in Spam or Adult Content Detection with very good results.

Note that each can deliver completely different results since they use completely different models.

Text can be represented in two separate ways. The first is as a bag of words, in which a document is represented as a set of words, together with their associated frequency in the

document. Such a Representation is essentially independent of the sequence of words in the collection. The second method is to represent text directly as strings, in which each document is a sequence of words. Most text classification methods use the bag-of-words representation because of its simplicity for classification purposes.

Text classifiers often do not use any kind of deep representation about language: often a document is represented as a bag of words. (A bag is like a set that allows repeating elements.) This is an extremely simple representation: it only knows which words are included in the document (and how many times each word occurs), and throws away the word order. Consider a document D, whose class is given by C. In the case of The Opportunities filtering there are two classes C = R (Course) and C = J (Job). We classify D as the class which has the highest posterior probability P(C|D), which can be re-expressed using Bayes' Theorem:

$P(C|D) = P(D|C)\ P(C)/\ P(D) \propto P(D|C)\ P(C)$ .

We shall look at two probabilistic models of documents, both of which represent documents as a bag of words, using the Naive Bayes assumption. Both models represent documents using feature vectors whose components correspond to word types. If we have a vocabulary V, containing |V| word types, then the feature vector dimension d=|V|, it means the number of different words in the vocabulary without duplicate the same word more than once.

Text Naïve Bayes Algorithm (Train)

1. Let V be the vocabulary of all words in the documents in D,
2. For each category ci $\in$ C
3. Let Di be the subset of documents in D in category ci
4. P(ci) = |Di| / |D|
5. Let Ti be the concatenation of all the documents in Di

6. Let ni be the total number of word occurrences in Ti
7. For each word wj ∈ V
8. Let nij be the number of occurrences of wj in Ti
9. Let P(wj | ci) = (nij + 1) / (ni + |V|)

Figure 2:23 Naïve Bayes Generative Model for Text

1. After getting the result and making the calculations for the probability, the following points should be considered:
    1.1. Multiplying lots of probabilities, which are between 0 and 1 by definition, can result in floating-point underflow.
    1.2. Since log(*xy*) = log(*x*) + log(*y*), it is better to perform all computations by summing logs of probabilities rather than multiplying probabilities.
    1.3. Class with highest final un-normalized log probability score is still the most probable.
    1.4. Classification results of Naïve Bayes (the class with maximum posterior probability) are usually fairly accurate.
    1.5. However, due to the inadequacy of the conditional independence assumption, the actual posterior-probability numerical estimates are not.
        - Output probabilities are generally very close to 0 or 1.

### 7.7. **Multinomial Naive Bayes Model**

This variation, as described by Manning et al (2008), estimates the conditional probability of a particular word/term/token given a class as the relative frequency of term t in documents belonging to class c:

$$P(t \mid c) = \frac{T_{ct}}{\sum_{t' \in V} T_{ct'}}$$

Thus, this variation takes into account the number of occurrences of term t in training documents from class c, including multiple occurrences.

Both the training and the testing algorithms are presented below in the form of pseudo code:

TRAINMULTINOMIALNB($\mathbb{C}, \mathbb{D}$)
1 $V \leftarrow$ EXTRACTVOCABULARY($\mathbb{D}$)
2 $N \leftarrow$ COUNTDOCS($\mathbb{D}$)
3 **for each** $c \in \mathbb{C}$
4 **do** $N_c \leftarrow$ COUNTDOCSINCLASS($\mathbb{D}, c$)
5 $prior[c] \leftarrow N_c / N$
6 $text_c \leftarrow$ CONCATENATETEXTOFALLDOCSINCLASS($\mathbb{D}, c$)
7 **for each** $t \in V$
8 **do** $T_{ct} \leftarrow$ COUNTTOKENSOFTERM($text_c, t$)
9 **for each** $t \in V$
10 **do** $condprob[t][c] \leftarrow \frac{T_{ct}+1}{\sum_{t'}(T_{ct'}+1)}$
11 **return** $V, prior, condprob$

APPLYMULTINOMIALNB($\mathbb{C}, V, prior, condprob, d$)
1 $W \leftarrow$ EXTRACTTOKENSFROMDOC($V, d$)
2 **for each** $c \in \mathbb{C}$
3 **do** $score[c] \leftarrow \log prior[c]$
4 **for each** $t \in W$
5 **do** $score[c] \mathrel{+}= \log condprob[t][c]$
6 **return** $\arg\max_{c \in \mathbb{C}} score[c]$

Figure 2.23 Naïve Bayes Algorithm

We will explain how Naïve Bayes Algorithm help us to find the matching process between the PALO, MOL and Jobs outcomes. In the following example, we have three learning outcomes for course (MLO), two jobs learning outcomes and achieved learning outcome for the learning (PALO).

Bayes theorem provides a way of calculating the posterior probability, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. Naive Bayes classifier assumes that the effect of the value of a predictor ($x$) on a given class ($c$) is independent of the values of other predictors. This assumption is called 'class conditional independence'.

| Doc ID | Document (Learning Outcome) | Class | W Count |
|---|---|---|---|
| 1 | Develop~~ing~~ Microsoft SQL Server 2012 | Course | 5 |
| 2 | Develop~~ing~~ mobile application~~son~~ the android | Course | 4 |
| 3 | learning c#~~for~~ begin~~nersand~~ SQL Server | Course | 5 |
| 4 | Senior .NET Developer | Job | 4 |
| 5 | Oracle Database Administrator | Job | 3 |
| PALO | 5 Year~~s'~~ experience ~~in~~ Microsoft SQL Server 2012 Database | ?? | 10 |

Table 7 : Classification and Naïve Bayes – Example

We will apply the Naïve Bayes algorithm for the above example to check the personal achieved learning outcomes belong for which one **job** or **course**.

What is the probability that a person has learning opportunities or job opportunities?

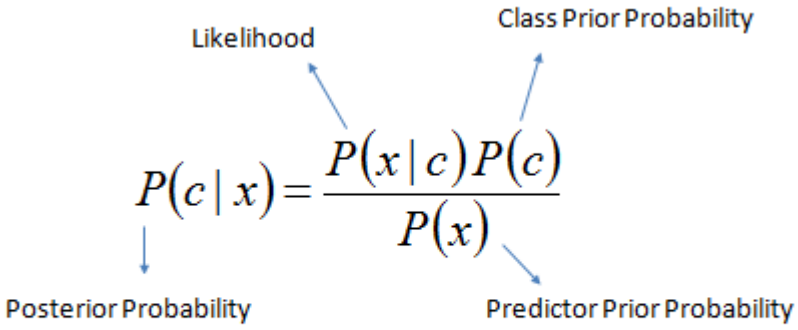
***Person has Course if***
***P (Course | x i) > P (Jobs | x i),***
***Else classify person as Job.***

**Step-1:** Calculate the prior probabilities; what is the probability for each class.

P(class) =# of first class features in training data/# of all Class(all jobs and Courses documents ) in training data

**Step -2:** Calculate the **class conditional probabilities**, which means that the probabilities occurs for every words in **PALO** in every class (likelihood).

Before stating the matching, we need to apply many string manipulations process to get the parameters ready for matching.




$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

- $P(c|x)$ is the posterior probability of *class* (*target*) given *predictor* (*attribute*).
- $P(c)$ is the prior probability of *class*.
- $P(x|c)$ is the likelihood which is the probability of *predictor* given *class*.
- $P(x)$ is the prior probability of *predictor*.

Example:

The posterior probability can be calculated by first constructing a frequency table for each attribute against the target. Then, transforming the frequency tables to likelihood tables and finally use the Naive Bayesian equation to calculate the posterior probability for each class.

**The class with the highest posterior probability is the outcome of prediction**.

**Priors:**
$P(course) = \frac{3}{5}$
$P(Job) = \frac{2}{5}$

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(w \mid c) = \frac{count(w,c)+1}{count(c)+|V|}$$

**Developing Microsoft SQL Server 2012**

Developing mobile applications on the android

learning c# for beginners and SQL Server

5 Years' experience in Microsoft SQL Server 2012 Database

Conditional Probabilities: (Query(PALO) with Courses)

P(5|course) = (0+1) / (14+17) = 1/31 = 0.032

P(Year|course) = (0+1) / (14+17) = 1/31= 0.032

P(experience |course) = (0+1) / (14+17) = 1/31 = 0.032

P(Microsoft|course) = (1+1) / (14+17) = 2/31 = 0.064

P(SQL|course) = (2+1) / (14+17) = 3/31 = 0.096

P(Server|course) = (2+1) / (14+17) = 3/31 = 0.096

P(2012|course) = (1+1) / (14+17) = 2/31 = 0.064

P(Database|course) = (1+1) / (14+17) = 2/31 = 0.064

**Senior .NET Developer**

Oracle Database Administrator

5 Years' experience in Microsoft SQL Server 2012 Database

Conditional Probabilities: (Query(PALO) with Jobs)

P(5|Job) = (0+1) / (6+17) = 1/23 = 0.043

P(Year| Job) = (0+1) / (6+17) = 1/23 = 0. 043

P(experience | Job) = (0+1) / (6+17) = 1/23 = 0.043

P(Microsoft| Job) = (0+1) / (6+17) = 2/23 = 0.086

P(SQL| Job) = (0+1) / (6+17) = 3/23 = 0.13

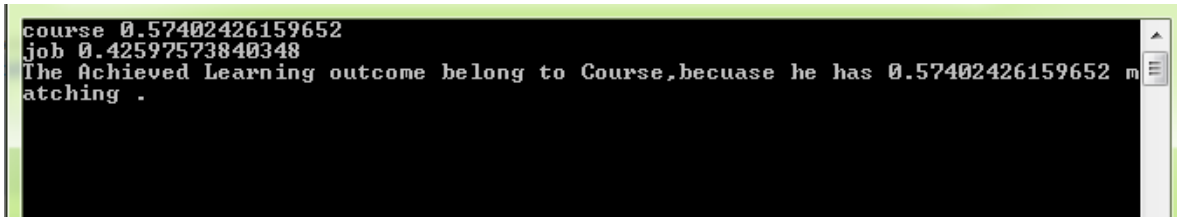
P(Server| Job) = (0+1) / (6+17) = 3/23 = 0.13

P(2012| Job) = (0+1) / (6+17) = 2/23 = 0.086

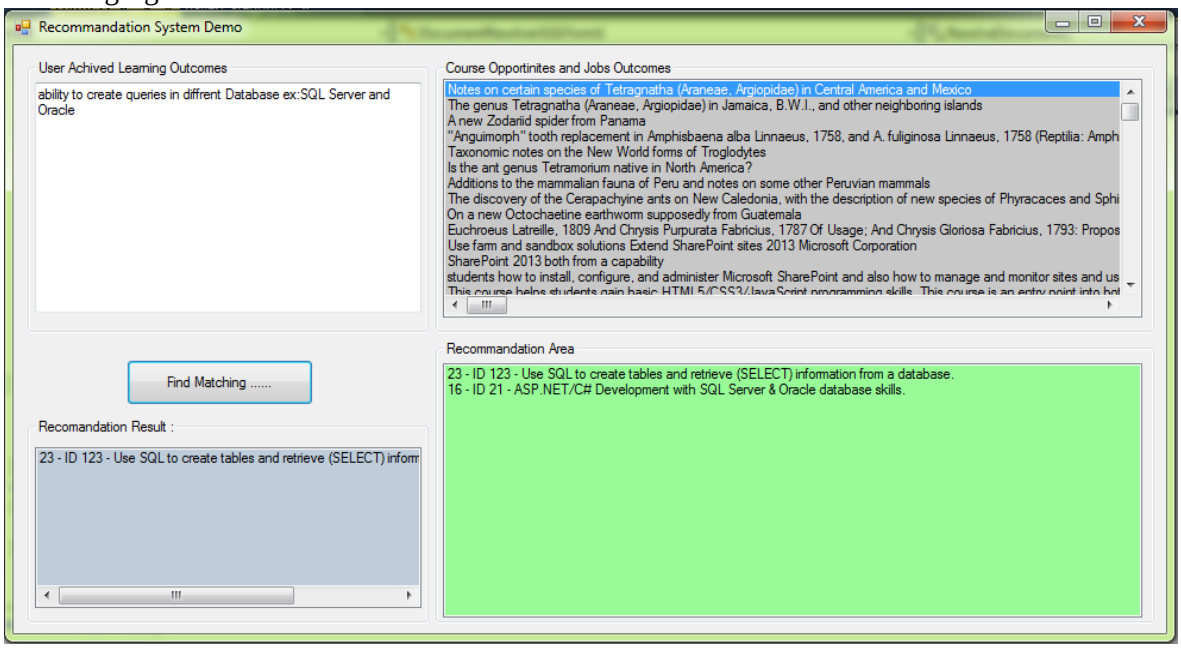
P(Database| Job) = (1+1) / (6+17) = 2/23 = 0.086

Now, we will apply this Example to Naïve Bayes algorithm like the following to check the matching Process between the outcomes:

```
staticList<Document> _trainSet = newList<Document>
        {
newDocument("course", "Develop Microsoft SQL Server 2012"),
newDocument("course", "Develop mobile application android"),
newDocument("course", "learn c# for begin SQL Server"),
newDocument("job", "Senior .NET Develop"),
newDocument("job", "Oracle Database Administrator")
        };

staticstringachieved_LO = "5 Year experience Microsoft SQL Server 2012 Database";


staticvoid Main(string[] args)
        {
var c = newClassifier(_trainCorpus);
varIs_belong_to_Course = c.IsInClassProbability("course", achieved_LO);
varIs_belong_to_Job = c.IsInClassProbability("job", achieved_LO);

Console.WriteLine("course " + Is_belong_to_Course);
Console.WriteLine("job " + Is_belong_to_Job);
if (Is_belong_to_Course>Is_belong_to_Job)
            {
Console.WriteLine("The Achieved Learning outcome belong to Course,becuase he has " + Is_belong_to_Course +
" matching .");
            }
else
            {
Console.WriteLine("The Achieved Learning outcome belong to Job " + Is_belong_to_Course);
            }
Console.ReadKey();
        }
```




Also, we will apply another example as desktop app to explain the matching result too. See the following figure:



# Chapter 3

## System and Design

### 1. Output and User Interface Design

#### 1.1. **Output Design:**

Although business information systems still provide most output as screen displays and printed matter, technology is having an enormous impact on how people communicate and obtain information. This trend is especially important to firms that use information technology to lower their costs, improve employee productivity, and communicate effectively with their customers.

##### 1.1.1. **Types of Output:**

Internet-Based Information Delivery**:** Millions of firms and universities use the Internet to reach new Training courses or Jobs around the world. Web designers must provide user-friendly screen interfaces that display output and accept input from Learners. (Forms)

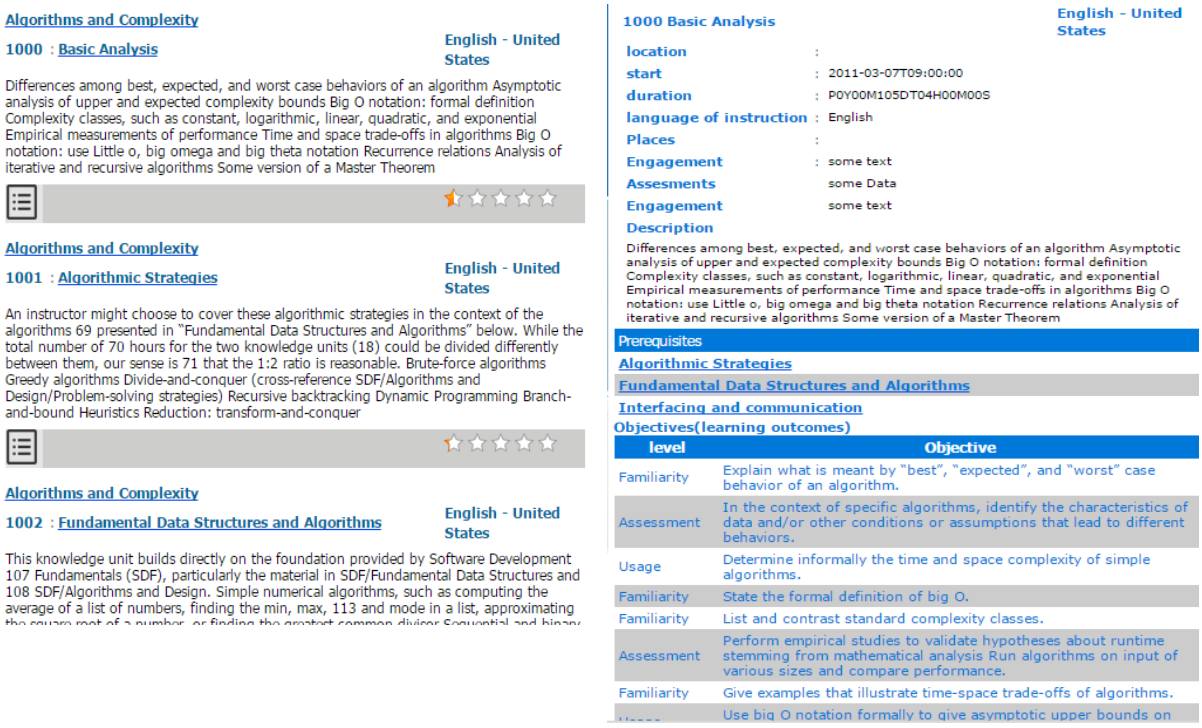



Figure 3:1 Learning Opportunities Outcomes


*Figure 3:2 Job Details*


*Figure 3:3 Achieved Outcomes and learning path*

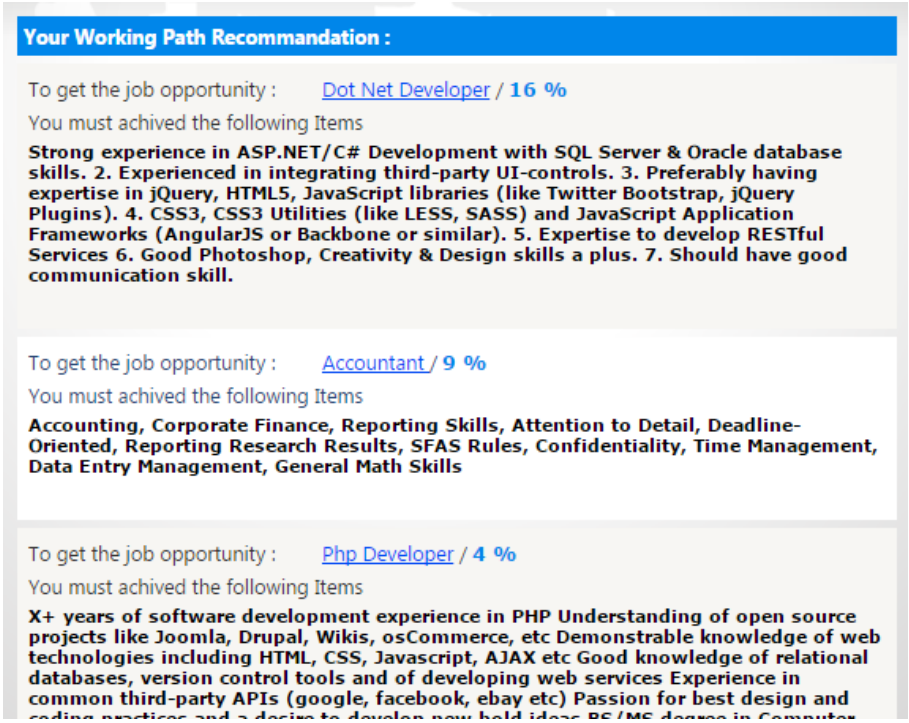



*Figure 3:4 Working Path*


*Figure 3:5 Related Items*

1.1.2. **E-MAIL:** E-mail is an essential means of internal and external business communication. Learners send and receive e-mail on local or wide area networks, including internet, interviews emails, latest news email, remember and activation password, activation links and others.

1.1.3. **Blogs and Widget:** Web-based logs, called **blogs**, are another form of Web-based output. Since blogs are journals written from a particular point of view, they do not only deliver facts to Web readers, but also provide opinions. Blogs are useful for posting news, reviewing current events, and promoting products. Display the Comments for learning outcome and display the voting result.


*Figure 3:6 Recommendation Area*

1.1.4. **Popup Screen:** A pop-up screen displays some important actions.

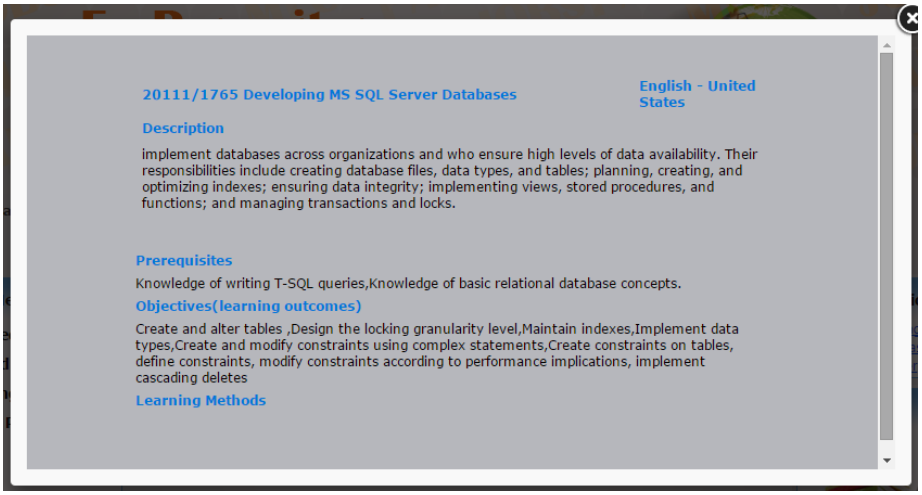



*Figure 3:7 Popup Course Details*

1.1.5. **SMS**: Send messages to learners in order to confirm the successful creation profile, an active profile or to send notifications when the system has new events or new courses or job.

1.1.6. **Printing Reports.**

## 1.2. User Interface Design

A **User Interface (UI)** describes how users interact with a computer system, and consists of all the hardware, software, **screens**, **menus**, functions, **output** and features that affect two-way communications between the user and the computer. Good user interface design is based on the following items: build an interface that is easy to learn and use, provide features that promote efficiency, make it easy for users to obtain help or correct errors, minimize input data problems, provide feedback to users, great an attractive layout and design and use familiar terms and images.

**User Interface Controls types:**

- Home Page:


*Figure 3:8 Home Page*

- Menus:

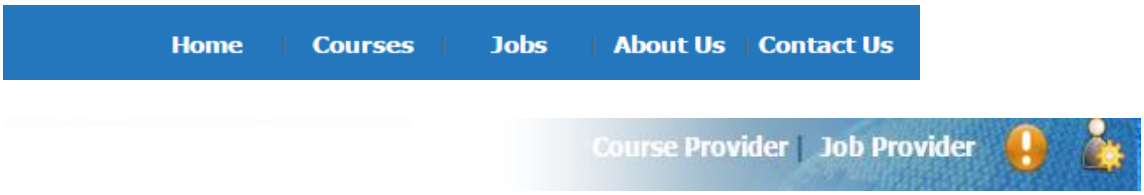
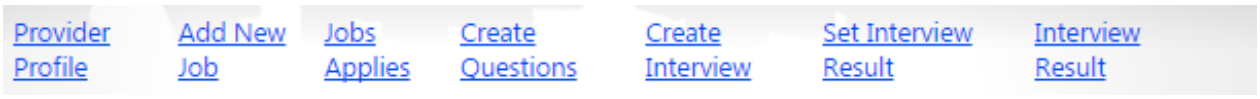



*Figure 3:9 Menus*

- Provider Main Menu:



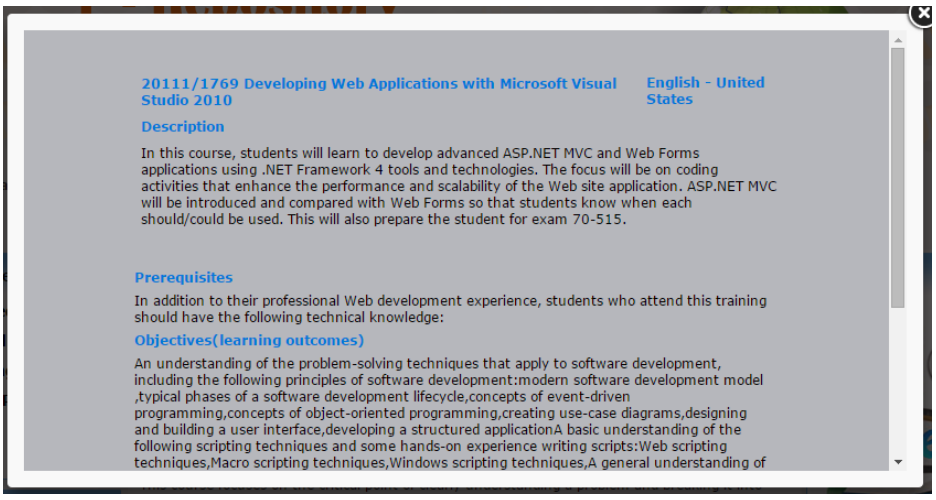

- Learning Provider Menu:

- Vertical Menus:

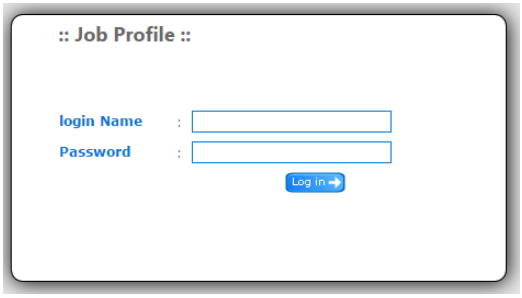

- Command Button:

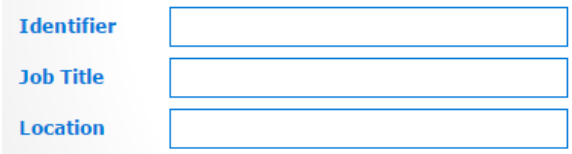

*Figure 3:10 Vertical Menus and Voting Bar*

- Dialog Box:

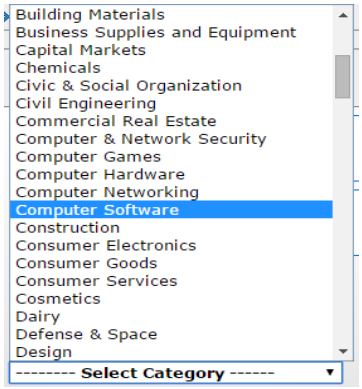


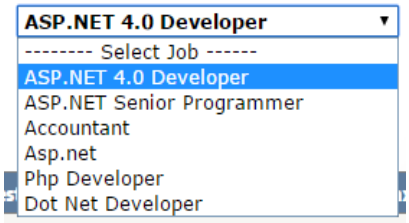

- Text Box:



- Drop-down List Box:

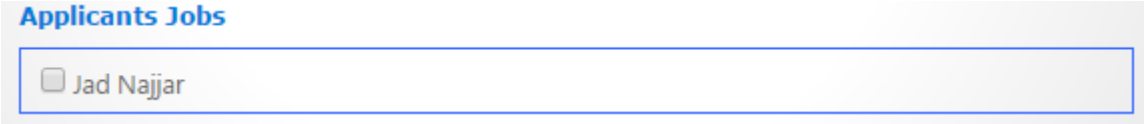



- Data Grid:

| Job Name | Question Types | Question Txt | Max.Score |
|---|---|---|---|
| ASP.NET 4.0 Developer | Motives | What Is ..... | 12 |
| ASP.NET 4.0 Developer | Marks | Where | 45 |
| ASP.NET 4.0 Developer | Motives | ttttt | 44 |
| ASP.NET 4.0 Developer | Goals | where is ..... | 10 |

- Check Box:




- Calendar Control:

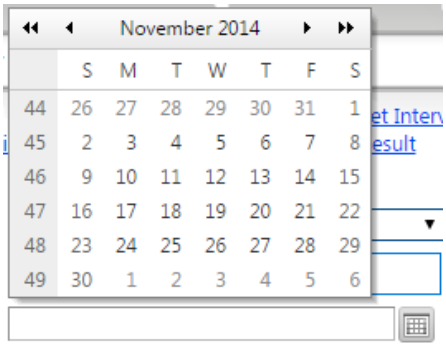



*Figure 3:11 Snapshot for the Recommendation System*

### 1.3. **Input Design**

The quality of the output is only as good as the quality of the input, most effective method of online data entry is from filling, and the input processes should be efficient, timely and logical; we have two method to insert the Date in the System.

1. **Batch Input:** Using **batch input**, data entry is usually performed on a specified time schedule, such as daily, weekly and monthly. We will use this method in Web services, rss reader input and atom feed parser, where this data comes from the Provider directly who exports the data to the system like learning opportunity provider, PALO Profile and Job Providers.
2. The second is Data Online Input, when the learner edits his profile; adds comments, adds/Updates Prerequisites or Objectives by using the keyboard buttons, **data entry.**

### 1.4. **Input Errors**

An effective way to reduce input errors is to reduce input volume. Using well designed data entry screens and using data validation checks can reduce such errors.

The main types of data validations are:

1. Compare Validator: Compares the value of one input control to the value of another input control or to a fixed value. This is used when we need to change the user password it needs to retype it to insure you insert the same password.
2. Custom Validator: Allows us to write a method to handle the validation of the value entered, and this validation is used when we need to create java script validation.

3. Range Validator: Checks that the user enters a value that falls between two values, like when the user specifies the user age between 20 to 35 who use the system.
4. Regular Expression Validator: Ensures that the value of an input control matches a specified pattern like the email format
5. Required Field Validator: Makes an input control a required field.

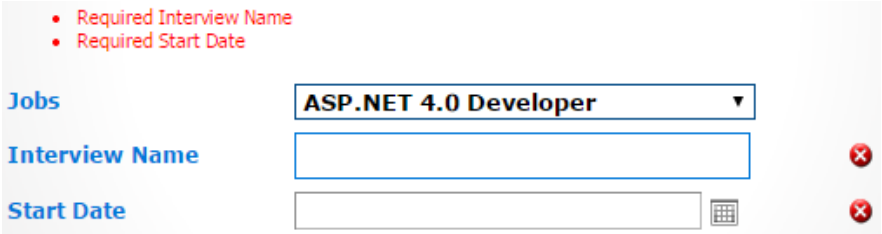



*Figure 3:12 Input Validation*

# 2. Data Design

## 2.1. **Data Structures**

Each file or table contains data about people, places, things or events that interact with the information system.

1. File-oriented System: a file-oriented approach to storage creates files in sets as needed when a business sells products or services. Each file operates independently from other files in storage. This means files don't share information with other files stored in the system.
2. File Processing System:
   2.1. File processing system stores data in separate computer files. File processing system is a system used to store and manage data that involves each department or area within an organization having its own set of files, often creating data redundancy and data isolation.

2.2. File processing can be efficient and cost- effective in certain situations.

2.3. Potential problems.

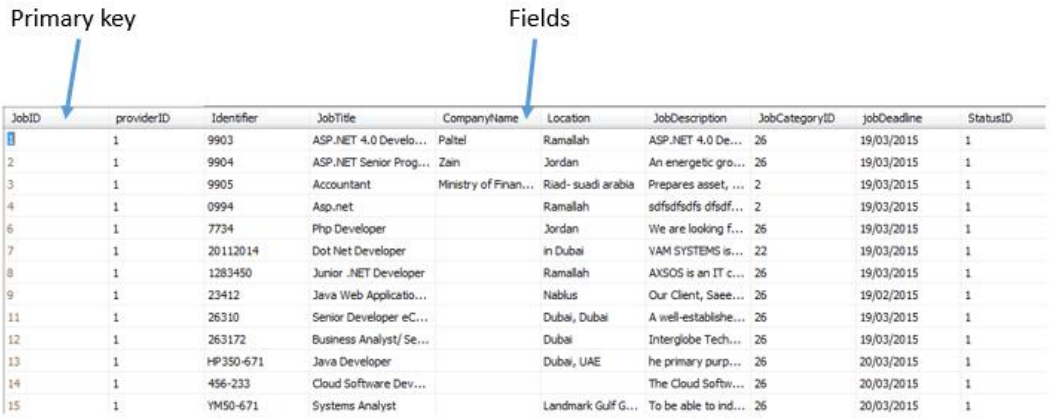
2.1.1. Data redundancy

Data redundancy in database means that some data fields are repeated in the database, and this data repetition may occur either if a field is repeated in two or more tables, or if the field is repeated within the table. Such repetition cases the database result in data inconsistency, decrease the efficiency of the database and increase the size of the database unnecessarily.

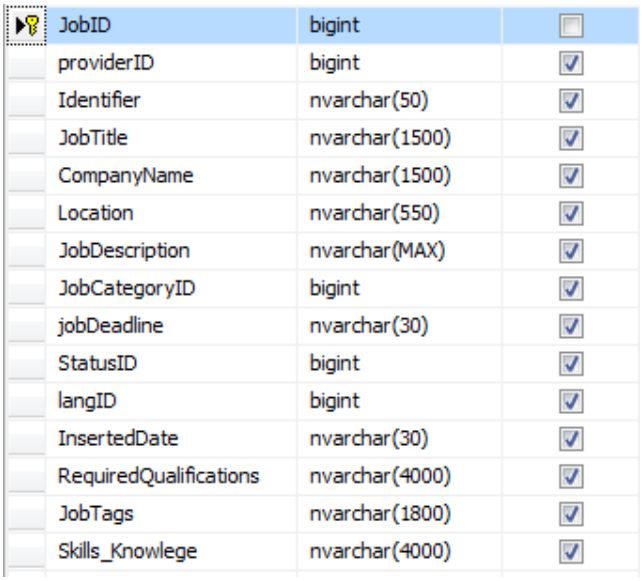
2.1.2. Data Integrity

2.1.3. Database System

A database management system (DBMS) is a collection of tools, features and interfaces that enables users to add, update, manage, access and analyze the contents of a database. It allows the users to create a database that has table(s) to store data like (job description) table, and each tables has many records like (job name, category, job dead line), every record has many fields(Job Name) and every field consists of many characters like (Asp.net).




| JobID | providerID | Identifier | JobTitle | CompanyName | Location | JobDescription | JobCategoryID | jobDeadline | StatusID |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 9903 | ASP.NET 4.0 Develo... | Paltel | Ramallah | ASP.NET 4.0 De... | 26 | 19/03/2015 | 1 |
| 2 | 1 | 9904 | ASP.NET Senior Prog... | Zain | Jordan | An energetic gro... | 26 | 19/03/2015 | 1 |
| 3 | 1 | 9905 | Accountant | Ministry of Finan... | Riad- suadi arabia | Prepares asset, ... | 2 | 19/03/2015 | 1 |
| 4 | 1 | 0994 | Asp.net | | Ramallah | sdfsdfsdfs dfsdf... | 2 | 19/03/2015 | 1 |
| 6 | 1 | 7734 | Php Developer | | Jordan | We are looking f... | 26 | 19/03/2015 | 1 |
| 7 | 1 | 20112014 | Dot Net Developer | | in Dubai | VAM SYSTEMS is... | 22 | 19/03/2015 | 1 |
| 8 | 1 | 1283450 | Junior .NET Developer | | Ramallah | AXSOS is an IT c... | 26 | 19/03/2015 | 1 |
| 9 | 1 | 23412 | Java Web Applicatio... | | Nablus | Our Client, Saee... | 26 | 19/02/2015 | 1 |
| 11 | 1 | 26310 | Senior Developer eC... | | Dubai, Dubai | A well-establishe... | 26 | 19/03/2015 | 1 |
| 12 | 1 | 263172 | Business Analyst/ Se... | | Dubai | Interglobe Tech... | 26 | 19/03/2015 | 1 |
| 13 | 1 | HP350-671 | Java Developer | | Dubai, UAE | he primary purp... | 26 | 20/03/2015 | 1 |
| 14 | 1 | 456-233 | Cloud Software Dev... | | | The Cloud Softw... | 26 | 20/03/2015 | 1 |
| 15 | 1 | YM50-671 | Systems Analyst | | Landmark Gulf G... | To be able to ind... | 26 | 20/03/2015 | 1 |



| | | |
|---|---|---|
| JobID | bigint | ☐ |
| providerID | bigint | ☑ |
| Identifier | nvarchar(50) | ☑ |
| JobTitle | nvarchar(1500) | ☑ |
| CompanyName | nvarchar(1500) | ☑ |
| Location | nvarchar(550) | ☑ |
| JobDescription | nvarchar(MAX) | ☑ |
| JobCategoryID | bigint | ☑ |
| jobDeadline | nvarchar(30) | ☑ |
| StatusID | bigint | ☑ |
| langID | bigint | ☑ |
| InsertedDate | nvarchar(30) | ☑ |
| RequiredQualifications | nvarchar(4000) | ☑ |
| JobTags | nvarchar(1800) | ☑ |
| Skills_Knowlege | nvarchar(4000) | ☑ |

*Figure 3:13 Table's Structures*

The main advantage of a DBMS is that it offers timely, interactive and flexible data access.

1. Scalability;
2. Economy of scale;
3. Enterprise- wide application –database administrator (DBA);
4. Controlled redundancy;
5. Data independence.

RDBMS: is the Relational Database Management System in which a database has many tables and has relations between these tables as shown in Fig 3.12.

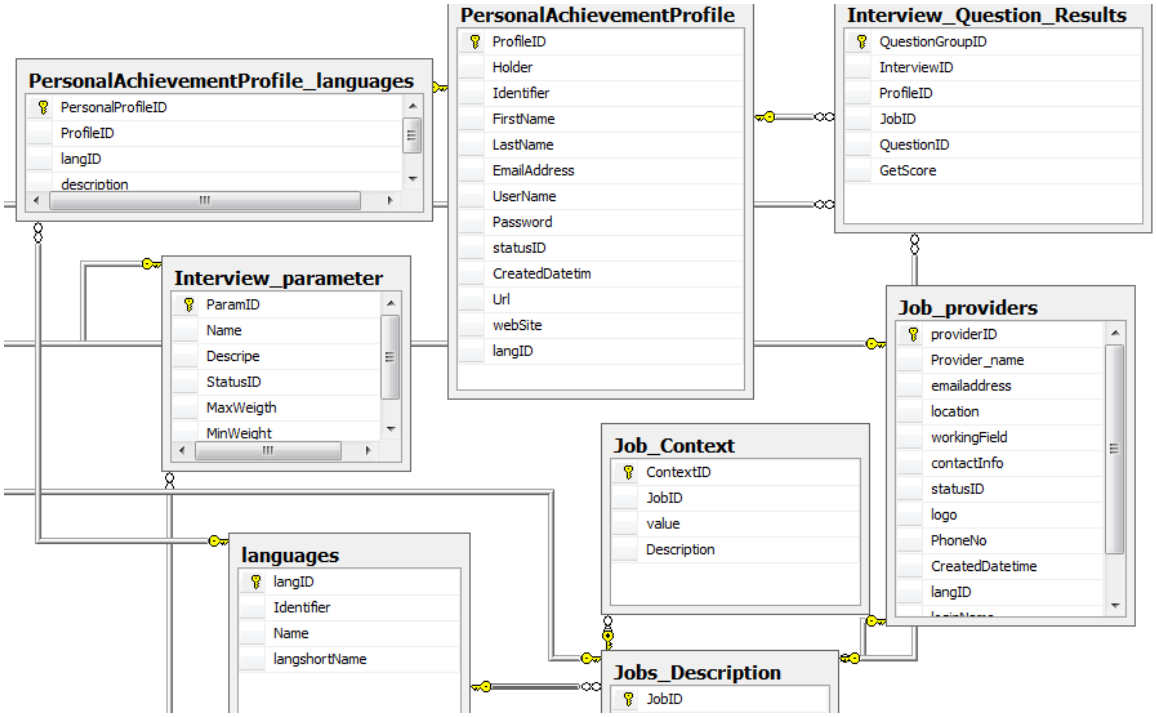



*Figure 3:14 Relational Database*

2.2. **DBMS Components**

2.3. **Database Engine:**

The Database Engine is the core service for storing, processing, and securing data. The Database Engine provides controlled access and rapid transaction processing to meet the requirements of the most demanding data consuming applications within your enterprise.

- **Data Dictionary:**

A Data Dictionary is a reserved space within a database which is used to store information about the database itself. A Data Dictionary is a set of tables and views which can only be read and never altered.

A Data Dictionary may contain information such as:

1. Database design information;
2. Stored SQL procedures;

3. User permissions;
4. User statistics;
5. Database process information;
6. Database growth statistics;
7. Database performance statistics.

- Query Processor:

The query processor accepts SQL syntax, selects a plan for executing the syntax, and then executes the chosen plan. The user or program interacts with the query processor, and the query processor in turn interacts with the storage engine. The query processor components include:

- DDL interpreter
- DML compiler
- Query evaluation engine

## 2.4. **Web-Based Database Design**

The Internet enables world wide access, using existing infrastructure and standard telecommunication protocols. Web-based design is not dependent on a specific combination of hardware or software. All that is required is a browser and an internet connection, where web browser provides a familiar interface that is user-friendly and easy to learn and Initial investment is relatively low, flexibility is high.

For creating web-based Database design we need focus on the following items:

1. **Internet Terminology Web browser**
    1. Web page;
    2. HTML (Hypertext Markup Language);

3. Tags;
4. Web Server;
5. Website;
6. Intranet;
7. Extranet;
8. Protocols;
9. Web-centric;
10. Clients;
11. Servers.

Connecting a database to the web: database must be connected to the Internet or intranet.

**Database Diagrams:**

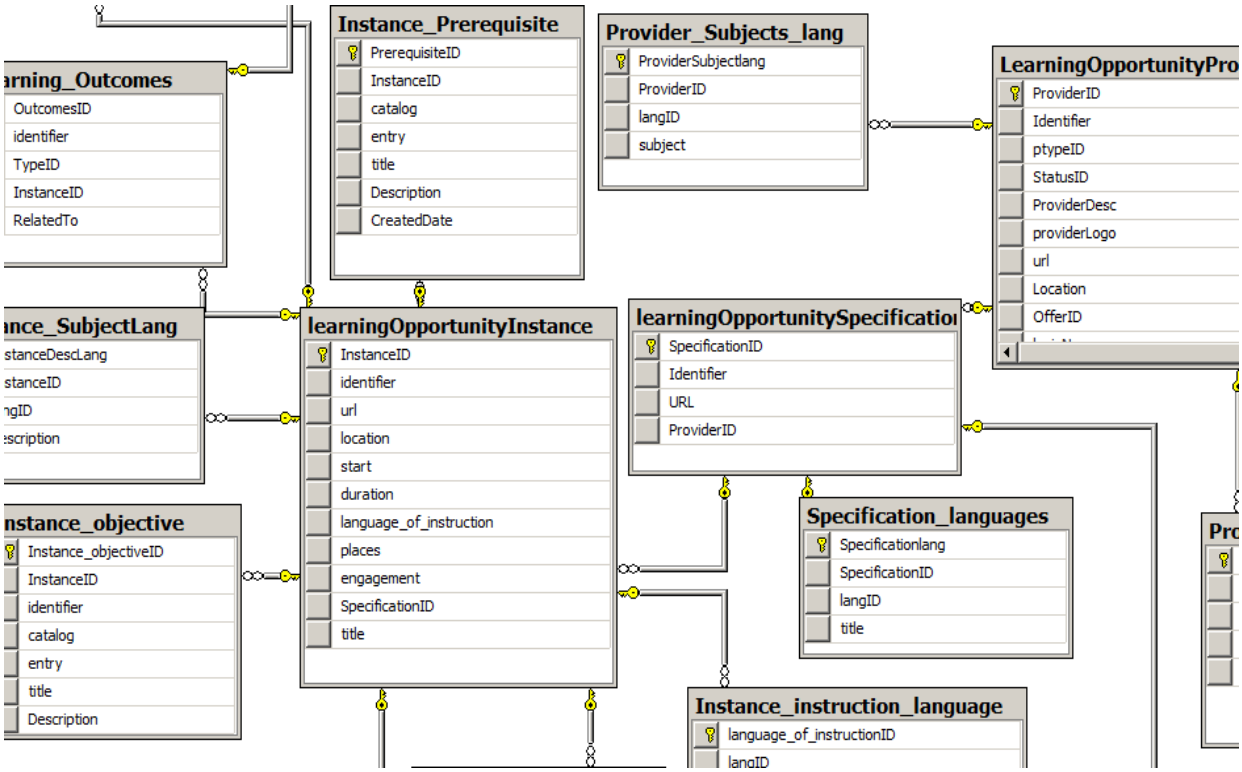



*Figure 3:15 MLO Database Diagram*

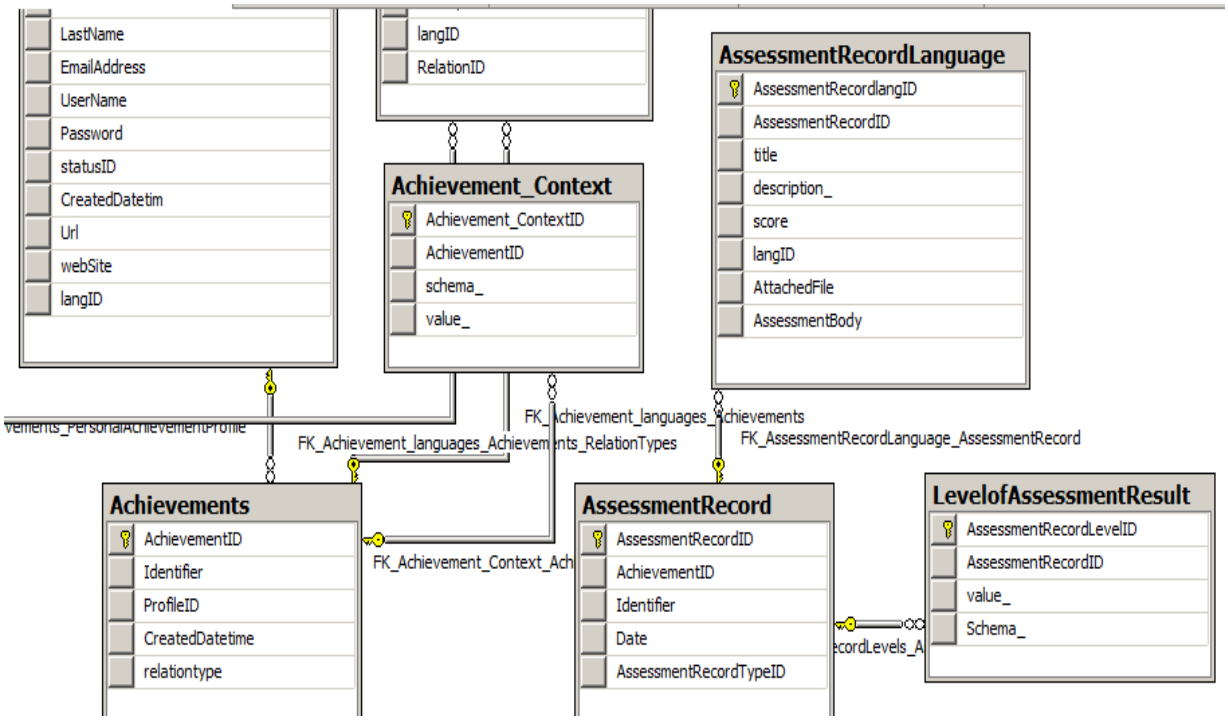



*Figure 3:16 PALO Database Diagram*

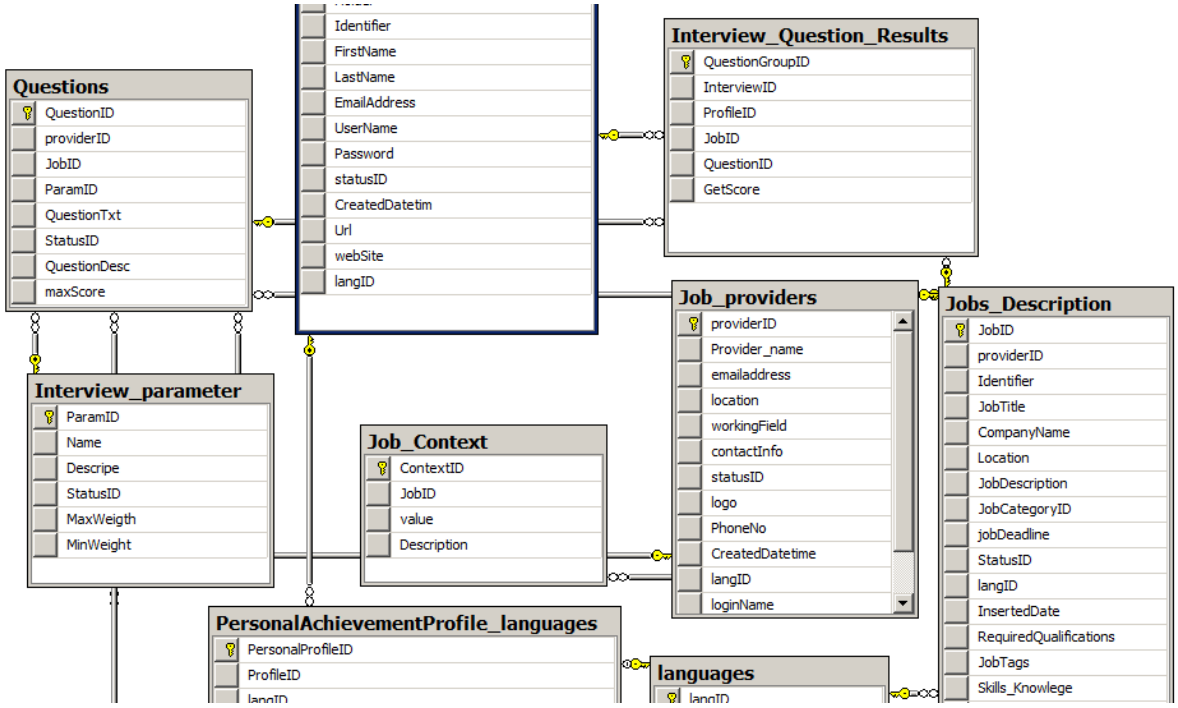



*Figure 3:17 Job Tables*

## 2.5. **System Architecture**

An effective system combines elements into an architecture, or design that is flexible, cost-effective, technically sound, and able to support the information needs of the business. The architecture require servers and clients.

The matching engine depends on three main components:

- Metadata Learning Opportunity (**MLO**): the importance of this component is to provide learning Opportunities from Providers who provide the learning like courses. We can read the data from the provider as xml file, web servers rss feed or atom feed, by creating the defined xml elements that form the attributes for the MLO Standard that we discussed in Section 1. The parser can read the elements for the Instance Object with all attributes and store the data in database, it consists of the following sections: provider information, instance Information, instance learning outcomes and prerequisites.

When the engine system is working, it can read the available learning opportunity outcomes from the database and apply many functions to prepare the data for matching.

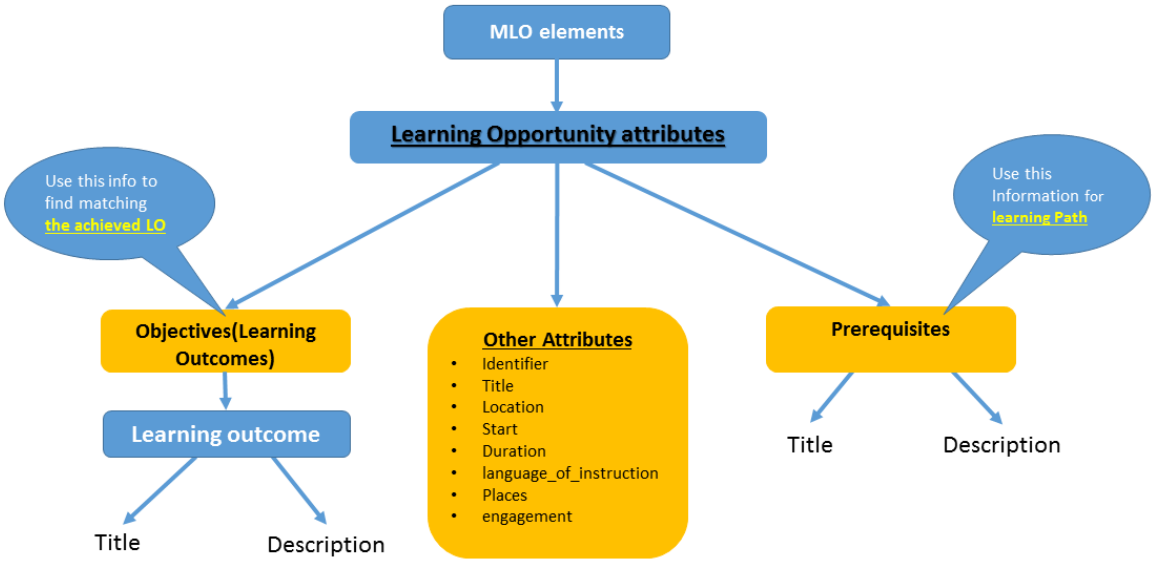



*Figure 3:18 MLO Attributes*

- Personal Achieved Learning Outcomes (**PALO**): this is the second component on matching engine that deals with the user profile elements, and this component has many define elements that allow us to read the values and store in Database. It consists of the following sections: personal achievement profile, achievement learning outcomes, intended outcomes and assessment records. After reading the data from the database, we apply many strings processing to be ready for matching.
- The third component is the job profile that consists of the information about the job like job title, description, skills, knowledge and prerequisites and others.

All these components work together to get recommendation percentage for every participant (Users) and drawing the learning path and career path for each one of them. The below figure (Fig. 3.19) explains the main components.

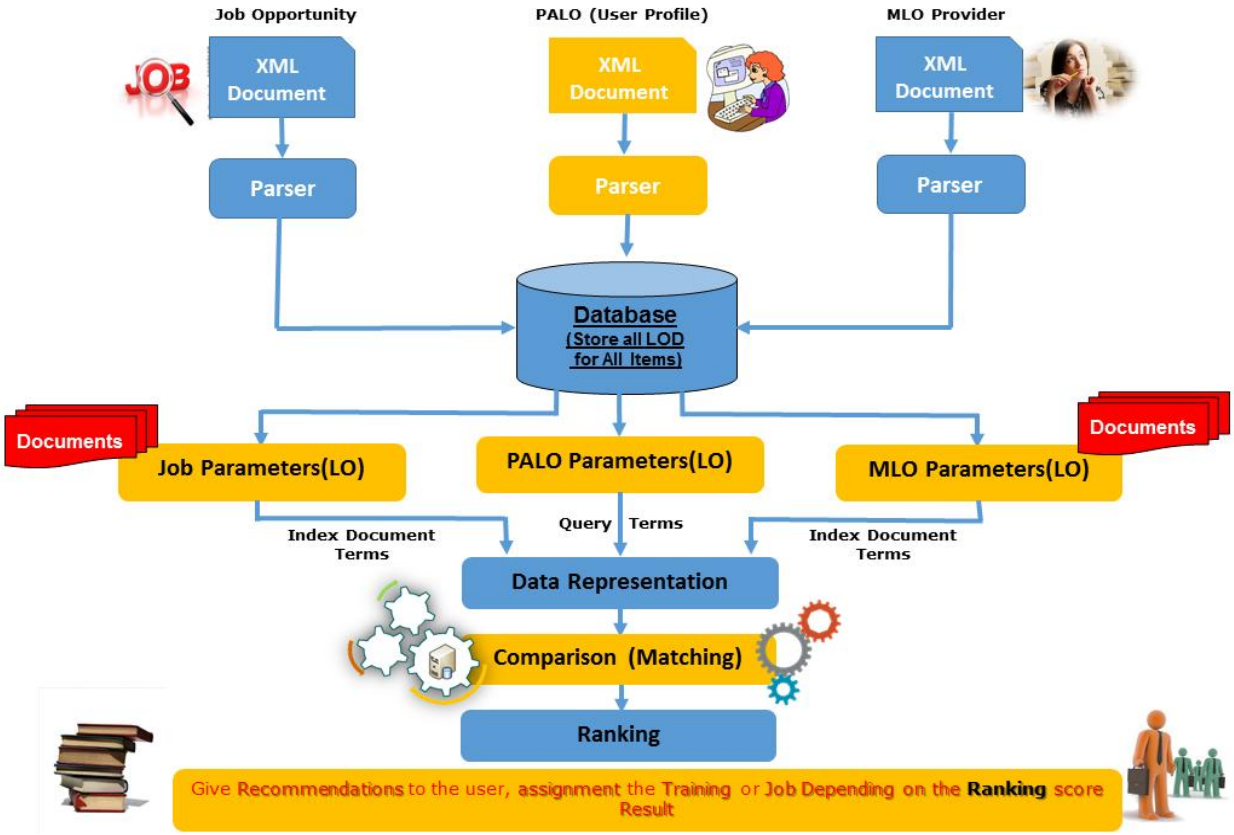



*Figure 3:19 System Architecture*

There are several communication channels between these components:

1. Learning Opportunities /Job Opportunities /Personal Profile
2. Candidates publish the data on the server and store it in database.
3. The server returns to Candidates to confirm or feedback confirmation.
4. The channel between the server and Database, when the query is executed or results are returned.
5. The channel between the clients (GUI) and server.
6. The channel between the server and Clients (GUI).

Depending on the previous channels, the most suitable approach to connect between these components as **Clients / Server Architecture**. Figure 3.20 displays the channels and clients / server Architecture.

The advantages for selecting the Clients / Server Architecture are:

1. Client/server systems enables companies to scale the system in a rapidly changing environment.
2. Client/server computing also allows companies to transfer applications from expensive mainframes to less expensive client platforms
3. Client/server systems reduce network load and improve response times

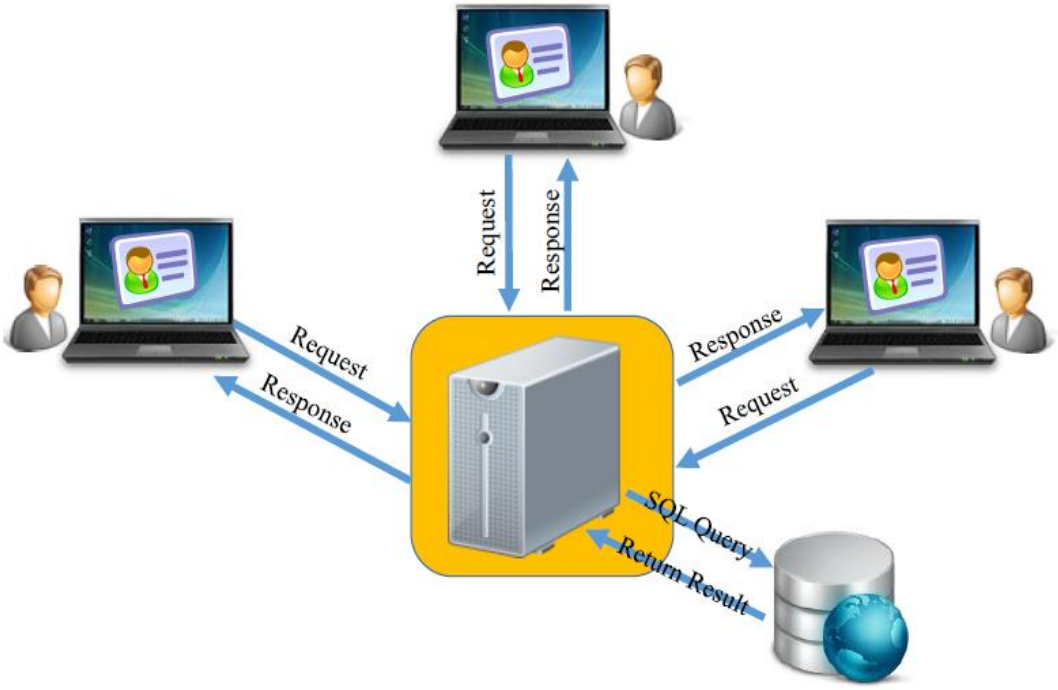



*Figure 3:20 Clients / Server Architecture*

# Chapter 4

## Implementation

The objective of the Implementation is to translate the design into program and code modules that will function properly.

1. **Technologies**

This chapter describes chosen technologies, important implementation decisions and other relevant implementation details.

The software platform chosen for the project is C#. It is a modern object-oriented Language that has a rich collection of core functions and many libraries we developed the application as web application using ASP.NET as Internet or intranet application, this can be executed with little or no additional requirements, and used SQL SERVER 2012Database.

The requirements that need to be used or executed for the application are:

1.1. **Install IIS**

IIS is a web server software package designed for Windows Server. It is used for hosting websites and other content on the Web.

Microsoft's Internet Information Services provides a graphical user interface (GUI) for managing websites and the associated users. It provides visual means of creating, configuring and publishing sites on the web. The IIS Manager tool allows web administrators to modify website options, such as default pages, error pages, logging settings, security settings and performance optimizations.

1.2. **Net Framework 4.0 or higher versions**

Net Framework is a programming infrastructure created by Microsoft for building, deploying, and running applications and services that use .NET technologies, such as desktop applications and Web services, web Application and mobile application. The .NET Framework contains three major parts:

- The Common Language Runtime;
- The Framework Class Library;
- ASP.NET;
- Support JavaScript.

2. **Tools**

1.2. **Visual Studio 2012/2013**

Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs for Microsoft Windows, as well as web sites, web applications and web services.


*Figure 4:1 Visual studio (IDE)*

1.3. **SQL Server Database**

SQL Server is a relational database management system (RDBMS) from Microsoft that's designed for the enterprise environment. SQL Server runs on T-SQL (Transact -SQL), a set of programming extensions from Sybase and Microsoft that

add several features to standard SQL, including transaction control, exception and error handling, row processing, and declared variables.


*Figure 4:2 SQL Server (Database)*

1.4. The following Tools have been used :

- SmartDraw 2010;
- Stylus.Studio.2009.XML.Enterprise;
- MIndMiniger;
- DbSchema;
- Visio.

## 3. Coding

Coding is the process of turning program logic into specific instructions that the computer system can execute.

Working from a specific design, we will use a Programming language to transform program logic into code statements.

```
protectedvoidPreparingData(longParParProfileID)
      {
         _dictionary.Clear();
stringvarItemQuery = "";
DataTabledtlearningOutcomes = newDataTable();
ArrayListInstanceObjectiveslist = newArrayList();
DataSetds_course = newDataSet();
DataTabledtCourses = newDataTable();
DataTabledtJob = newDataTable();
DataTabledtProfile = newDataTable();
ArrayListCourseInfo = newArrayList();
Common.BLL.Instance_languagesCourseObj = newCommon.BLL.Instance_languages();
Common.BLL.Instance_objectiveobjectiveObj = newCommon.BLL.Instance_objective();
Common.Model.Instance_objectiveInfoobjectiveInfo = newCommon.Model.Instance_objectiveInfo();

Common.BLL.Jobs_DescriptionJobObj = newCommon.BLL.Jobs_Description();
Common.BLL.PersonalAchievementProfileProfileObj = newCommon.BLL.PersonalAchievementProfile();
```

```
stringvarCurrentDate = System.DateTime.Now.ToString("dd/MM/yyyy");
ArrayListlistQuery = newArrayList();
dtCourses = CourseObj.GetDataSetAllInstance_languages().Tables[0];
dtJob = JobObj.GetAllDataSetJobs_Profile_SkillsOnly(1, varCurrentDate).Tables[0];
dtProfile = ProfileObj.GetDataSetProviderInfo_ForMatching_New(ParParProfileID).Tables[0];
MatchingJob_CourseelementObj = newMatchingJob_Course();          dtlearningOutcomes =
objectiveObj.GetDataSetAllInstance_objectiveOnly().Tables[0];
for (inti = 0; i<dtlearningOutcomes.Rows.Count; i++)
        {
          _dictionary.Add(dtlearningOutcomes.Rows[i].ItemArray[0].ToString() + "," +
dtlearningOutcomes.Rows[i].ItemArray[2].ToString() + "," +
dtlearningOutcomes.Rows[i].ItemArray[1].ToString(),
dtlearningOutcomes.Rows[i].ItemArray[5].ToString());
        }
for (int z = 0; z <dtJob.Rows.Count; z++) // Job Profile
        {
          _dictionary.Add("Jobs" + "," + dtJob.Rows[z].ItemArray[0].ToString(),
dtJob.Rows[z].ItemArray[1].ToString());
        }
for (int j = 0; j <dtProfile.Rows.Count; j++) // PALO Profile
        {
varItemQuery = "";
varItemQuery = dtProfile.Rows[j].ItemArray[2].ToString() + " " +
dtProfile.Rows[j].ItemArray[3].ToString();
ResolveDocument(varItemQuery.TrimEnd());
        }
DisplayRecommandationItems(list);
      }

doubleminScore = 0;
varfilteredResults = (from result in results whereresult.Score>= minScoreselect result);
// Display the results
foreach (ResolutionResult result infilteredResults)
           {
if (result.Score< 80)
              {
string output = string.Format("{0} , {1} , {2} , {3}", result.Score.ToString(), result.Key.Split(',')[0],
result.Key.Split(',')[1], result.Document);
list.Add(output);
              }
           }
list.Sort();

if (varItem[1].Trim().ToLower() == "course")
                    {
if (!listcourseID.Contains(varItem[2]))
                       {
ObjInstance = newCommon.BLL.Instance_languages();
InfoInstance = newCommon.Model.Instance_languagesInfo();
varCourseID = long.Parse(varItem[2]);
InfoInstance.InstanceID.InstanceID = varCourseID;
InfoInstance = ObjInstance.GetInstance_languagesById(InfoInstance);
courseList.Add(InfoInstance);
// Area01.InnerHtml += "" + varItem[1] + " :<a href='Course_details.aspx?InstanceID=" +
InfoInstance.InstanceID.InstanceID + "'>" + InfoInstance.title + " </a>" + "\n";
listcourseID.Add(varItem[2]);
                           x = x + 1;
                       }
                    }
elseif (varItem[1].Trim().ToLower() == "Jobs")
                    {
```

```
if (!listJobID.Contains(varItem[2]))
                    {
JobObj = newCommon.BLL.Jobs_Description();
JobInfo = newCommon.Model.Jobs_DescriptionInfo();
varJobID = long.Parse(varItem[0]);
JobInfo = JobObj.GetJobs_DescriptionById(varJobID);
listJobID.Add(varItem[2]);
jobsList.Add(listJobID);
//  Area01.InnerHtml += "" + varItem[1] + " : <a href='job_details.aspx?jobID=" + JobInfo.JobID + "'>" +
JobInfo.JobTitle + " </a>" + "\n";
                        x = x + 1;
                    }
                }
```

*Figure 4:3 Sample of Matching Engine Code*


*Figure 4:4 Matching Engine System Home Page*

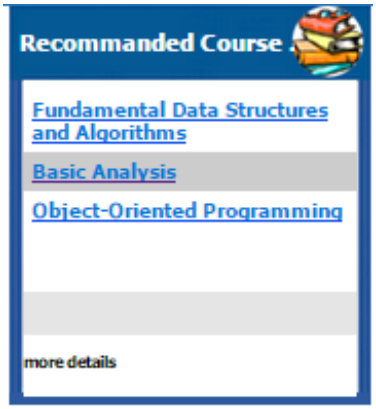



*Figure 4:5 Recommendation System Area*

# Chapter 5

## Evaluation and Testing

This section describes the approach used to define users' requirements and to evaluate the system. The first step was conducting interviews with two expert in order to get feedback from the recruitment company. The first interviewee is from Jobs.ps his profile is:

*(Amjad Hamarsheh is the founder of Jobs.ps, the first online job board company in Palestine based in Ramallah. Since 2009, Amjad is in charge of managing the company, developing business models, building new services, and acquiring new customers. Prior to founding jobs.ps, Amjad worked as IT trainer, focused on IT project management, ITIL. Amjad holds a Bachelor's Degree in MIS from the Arab American University)*

The second interviewee is the Head of Human Resources Section at Hadara Company. The interviewee's profile as follows:

*(Amir Helal is the Head of the Human Resources Section at Hadara. He is responsible for payroll administration, benefits and compensation, training & development, recruitment and selection, employee relations, reporting to the HR Manager, ensuring that all company policies and procedures are up to date and in-line with current employment laws, recruiting staff by preparing job descriptions and job adverts, deciding on the best ways to advertise, and planning and development of on-the-job training programs for line employees and career development programs for managers with a focus on retaining existing talent and reducing employee turnover).*

We to get the feedback about the course from Expert Turnkey Solution. The second step was conducting interviews to get and find out the requirements and feedback of end-users,

who in our case are referred to as the “learners”. Other sections in this Thesis provide more details about the evaluation process.

We conducted the interviews to analyze users’ requirements. Each group of users has its own concerns and expectations from the system. In order to identify user needs, interview with expert users have been conducted. The interviews aim at finding the user requirements and discovering probable problems which users face without matching the results. A usability test was performed, this step was implemented by conducting an interview to get requirements before the implementation phase of our system from the end users, and we collected feedback from interviews conducted with five participants (learners). Participants were selected as representatives of the intended user community.

- Introduction

A usability test is conducted in order to determine the usability of the recommendation result by measuring different metrics of usability. We have focused on the ease to learn, ease to use, user satisfaction, no errors and effective metrics

We prepared for the testing session by preparing a laptop on which the system was installed and running, Camtasia recording software, a questioner, a facilitator and note taker with the required tasks, formal consent form and a camera. The session captured the learning outcomes, course opportunity learning outcomes, and Job Opportunity learning outcomes each participant has achieved.

- Executive Summary

The test was conducted at the testing lab in Hadara Company. The user were selected randomly form different fields of specialization. We coordinated with the users one week prior to the testing. The Testing Session was conducted on Sunday, the 4$^{th}$ of December 2016, at 11:00am, with approximately 15-20 minutes for each user.

Our purpose was to evaluate our Matching result in order to measure practical it is through the usability measurement technique we mentioned before. In addition, we want to test for any problems/fault in the system functionality design, and repair them.

**Case -1**

Majd Archived LOC:

- Developing Web Applications using .Net and Angular js;
- Good experience in Web service;
- Implement basic numerical algorithms;
- implement common quadratic and O(N log N);
- Explain how binary search tree operations;
- Describe the heap property of priority queues.

Course Opportunity - Fundamental Data Structures and Algorithms:

| **1002 Fundamental Data Structures and Algorithms** | **English - United States** |
|---|---|

| Prerequisites |
|---|
| **Basic Analysis** |
| **Algorithmic Strategies** |

**Objectives(learning outcomes)**

| level | Objective |
|---|---|
| Usage | Implement basic numerical algorithms. |
| Assessment | Be able to implement common quadratic and O(N log N) sorting algorithms. |
| Familiarity | Describe the implementation of hash tables, including collision avoidance and resolution. |
| Familiarity | Discuss the runtime and memory efficiency of principal algorithms for sorting, searching, and hashing. |
| Familiarity | Discuss factors other than computational efficiency that influence the choice of algorithms, such as programming time, maintainability, and the use of application-specific patterns in the input data. |
| Familiarity | Explain how tree balance affects the efficiency of various binary search tree operations. |
| Usage | Solve problems using fundamental graph algorithms, including depth-first and breadth-first search. |
| Assessment | Demonstrate the ability to evaluate algorithms, to select from a range of possible |

| | |
|---|---|
| | options, to provide justification for that selection, and to implement the algorithm in a particular context. |
| Familiarity | Describe the heap property and the use of heaps as an implementation of priority queues. |
| Usage | Solve problems using graph algorithms, including single-source and all-pairs shortest paths, and at least one minimum spanning tree algorithm. |
| Usage | Trace and/or implement a string-matching algorithm. |

Job Opportunity - Front End Web Developer:

Responsibilities

-Front End Development

-Produce UI Mockups, Storyboards, Wireframes, User stories, interactive -Prototypes ( HTML5, CSS3, Bootstrap, JavaScript)

-Web Services Integration

-JQuery

-React and Angular

Education

Bachelor's Degree in Computer Science, Information Technology or related field

Required Skills and Qualifications

- Good Communication Skills
- Excellent problem solving skills
- Works well in teams
- Good organizational skills
- Consistently proactive
- Good experience in Web service and WCF
- Object Oriented Programming (OOP) concepts and design
- Experience developing cross-platform applications
- Web Application Development using Microsoft .NET Framework 4

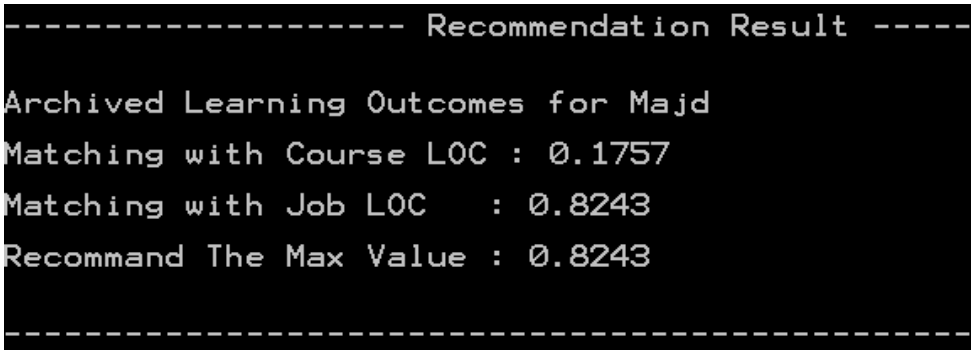



*Figure 5:1 Majd Matching Result*

**Case-2**

Taim Archived LOC:

- Good knowledge of web technologies;
- Good experience in JQuery and Ajxa;
- Good Experience with git and Scrum is a plus;
- Implement Class inheritance and polymorphism;
- Good Experience with Object Oriented Programming Language such as C#, Java, C++;
- Correctly reason about control flow in a program using dynamic dispatch;
- Good knowledge in c# and SQL Server 2014;
- Define and use iterators and other operations on aggregates;
- Design and implement a class.

Course Opportunity - Object-Oriented Programming:

**20111/1643 Object-Oriented Programming** **English - United States**

**Description**

- Object-oriented design
- Decomposition into objects carrying state and having behavior
- Class-hierarchy design for modeling
- Definition of classes: fields, methods, and constructors
- Subclasses, inheritance, and method overriding
- Dynamic dispatch: definition of method-call

| Prerequisites |
| --- |
| **Basic Analysis** |
| **Algorithmic Strategies** |
| **Fundamental Data Structures and Algorithms** |

**Objectives(learning outcomes)**

| level | Objective |
|---|---|
| Usage | Design and implement a class. |
| Usage | Use sub classing to design simple class hierarchies that allow code to be reused for distinct subclasses. |
| Usage | Correctly reason about control flow in a program using dynamic dispatch. |
| Assessment | Compare and contrast (1) the procedural/functional approach—defining a function for each operation with the function body providing a case for each data variant—and (2) the object-oriented approach—defining a class for each data variant with the class definition providing a method for each operation Understand both as defining a matrix of operations and variants. |
| Familiarity | Explain the relationship between object-oriented inheritance (code-sharing and overriding) and subtyping (the idea of a subtype being usable in a context that expects the super type). |
| Usage | Use object-oriented encapsulation mechanisms such as interfaces and private members. |
| Usage | Define and use iterators and other operations on aggregates, including operations that take functions as arguments, in multiple programming languages, selecting the most natural idioms for each language. |

Job Opportunity - Database System Developer:

Vacancy reference no.: VAC-0153

Position: Database System Developer

Place of performance: Gaza, Gaza

Contract duration: up to 4 months

Starting date: ASAP

ESSENTIAL EXPERIENCE

Education:

- Bachelor's degree in computer science, computer engineering or relevant discipline.

Key competencies:

- Intimate knowledge of web technologies (HTTP, HTML, CSS, JavaScript)
- Experience with any modern Object Oriented Programming Language such as C#, Java, C++, Python (we primarily use C#)
- Experience with any relational database system and SQL with good understanding of query performance optimization and data base normalization.
- Good understanding of what makes a good quality code and what is a code smell.

- Experience with git is a plus.
- Experience in networking, including setting up and supporting office IT infrastructure (i.e. configuring Active Directory, fixing printer issues, resolving internet connectivity issues etc.) is a plus.
- Good communication skill (spoken and written) in both Arabic and English..
- Personal interest, commitment, efficiency, flexibility and willingness to work in a highly motivated team of professionals.
- Able to work efficiently with minimal direction.
- Integrity and strong work ethics, prepared to work additional hours to ensure timely delivery of project objectives.
- Strong release, change and configuration management experience. -Excellent diagnostic, debugging and problem solving skills.

Other relevant information:

- Portfolio demonstrating past works and personal projects.
- Intimate knowledge of web technologies (HTTP, HTML, CSS, JavaScript)
- Experience with any modern Object Oriented Programming Language such as C#, Java, C++, Python (we primarily use C#)
- Experience with any relational database system and SQL with good understanding of query performance optimization and data base normalization.
- Strong release change and configuration management experience.
- Excellent diagnostic, debugging and problem solving skills.
- Contributor to open source projects.

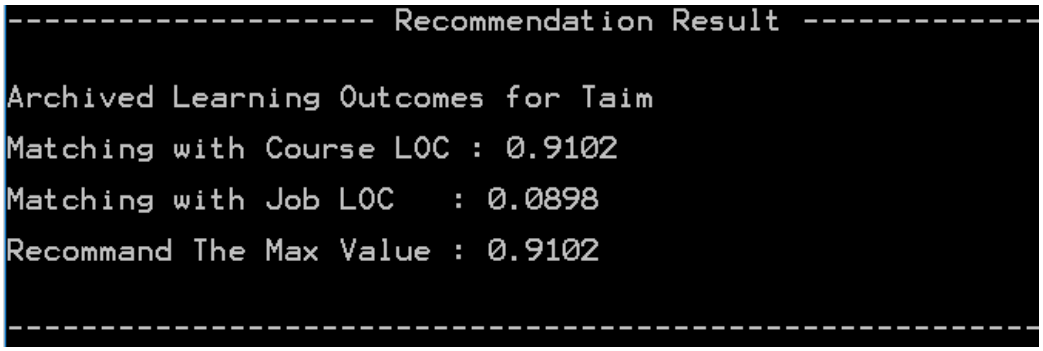



*Figure 5:2 Taim Matching Result*

**Case -3**

Alma Archived LOC:

- Professional certificate in database administration (MS, Oracle, MySQL) Proven experience in Database Performance Tuning
- Use a declarative query language to elicit information from a database
- Good Experience with Object Oriented Programming Language such as C#, Java, C++
- Presenting Technical Information, Quality Focus
- Describe major approaches to storing and processing large volumes of data
- Ability to Create Relational Database
- Define and use iterators and other operations on aggregates
- Familiarity with SQL Server Data Tools
- Database Management, Data Maintenance, Operating Systems, Information Security Policies

Course Opportunity - Database Systems:

**20111/1643 Database Systems**

**Description**

- Approaches to and evolution of database systems
- Components of database systems
- Design of core DBMS functions (e.g., query mechanisms, transaction management, buffer management, access methods)
- Database architecture and data independence
- Use of a declarative query language
- Systems supporting structured and/or stream content

| Prerequisites | |
|---|---|
| **Objectives(learning outcomes)** | |
| **level** | **Objective** |
| Familiarity | Explain the characteristics that distinguish the database approach from the approach of programming with data files. |
| Familiarity | Describe the most common designs for core database system components including the query optimizer, query executor, storage manager, access methods, and transaction processor. |
| Familiarity | Cite the basic goals, functions, and models of database systems. |
| Familiarity | Describe the components of a database system and give examples of their use. |
| Familiarity | Identify major DBMS functions and describe their role in a database system. |
| Familiarity | Explain the concept of data independence and its importance in a database system. |

| | |
|---|---|
| Usage | Use a declarative query language to elicit information from a database. |
| Familiarity | Describe facilities that databases provide supporting structures and/or stream (sequence) data, eg, text. |
| Familiarity | Describe major approaches to storing and processing large volumes of data. |

## Job Opportunity - SQL Database Administrator

**Responsibilities**

- Database Administration in SQL Server (2012 and newer)
- Ensure platform availability and viability.
- Identify and mitigate risk, affecting availability or performance.
- Detect, troubleshoot and resolve:
    - SQL Server related CPU, memory, I/O, disk space and other resource contention
    - Issues with database integrity, performance, blocking and deadlocking, connectivity or security
- Performance tune and optimize queries, using Performance Monitor, SQL Profiler and other related monitoring and troubleshooting tools
- Review, test and deploy change scripts for managed code and custom objects
- Manage and configure monitoring processes that are comprised of third party or custom components.
- Ensure all database platforms received a backed up in a manner that meets the Recovery Point Objective (RPO) and Recovery Time Objective (RTO).
- Perform database shrink operations, DBCC commands, clustering, database mirroring, replication
- Implement operational automation
- Windows server, security delegation, SPNs, storage components
- Documenting processes and procedures (creating KBs, runbooks, topology)
- SQL Database Operational support to tech users
- Remain informed of emerging technologies directly within or tangential to database technologies.

**Skills**

- Strong, demonstrated competency with TSQL and command level administration
- Experience with all aspects of SQL Server administration, including work with:
    - SQL Server Service
    - SQL Server Agent
    - SQL Server Reporting Services
    - SQL Server Integration Services
    - SQL Server Analysis Services
- Familiarity with SQL Server Data Tools
- Indexes, index management, integrity checks, configuration, patching.
- For a role in database management, employers will be looking for you to have the following:

- Strong analytical and organizational skills
- Eye for detail and accuracy
- Understanding of structured query language (SQL)
- Knowledge of 'relational database management systems' (RDBMS), 'object oriented database management systems' (OODBMS) and XML database management systems
- Experience with their database software/web applications
- The ability to work quickly, under pressure and to deadlines
- Up-to-date knowledge of technology and the Data Protection Act
- Ability to work well in a fast paced environment, where the technology is constantly changing

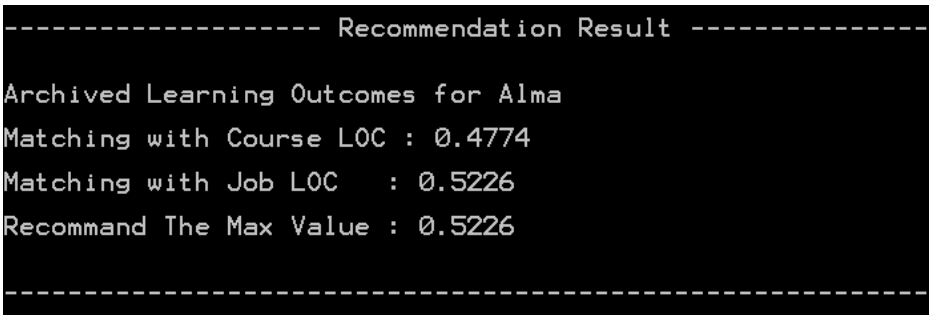



Figure 5:3 Alma Matching Result

After the completion of the testing results we get the following results for each case:

| Case # | Archived LOC | Course | Job | Recommend |
|---|---|---|---|---|
| **Case 1** | Majd | 0.1757 | 0.8245 | Job |
| **Case 2** | TAIM | 0.9102 | 0.0898 | Course |
| **Case 3** | ALMA | 0.4774 | 0.5226 | Job |
| | | | | |
| | | | | |

Table 8 : Matching Results for All Cases

## Conclusion and Future Work

In this paper we have presented concepts and solutions that were evaluated in real educational environments, which enable learners or Job Seekers to find the best recommendation based their achieved learning outcomes.

Our main goal is to facilitate the wider adoption and use of outcome based education and Jobs by educational institutions and recruitments companies.

In conclusion, the purpose of developing this system is to provide the below:

- The system will help recruitments companies in matching the best candidate for the best job profile.
- The system will enable universities to build dynamic profiles for their graduating students.
- Data matching can be used for conducting comparison between competing learners and employers, and for increasing the availability of detailed descriptions of learning opportunities by higher educational institutions.
- Investigating trust issues related to recognition of achieved learning outcomes, e.g. how to verify authenticity of assessment records or how to verify assessment bodies that issued assessment records and trust them that they really assessed achievements.

## Future work

In this section, we provide the future directions of this research. This includes the following:

1. Applying semantic search to facilitate the selection of the best matching results between candidates.

2. Finding the competitive points between candidates in order to enable Companies to select the best choice.
3. Improving outcome based search for learning designs, learning opportunities, job description and assessments using mechanisms that go beyond keywords search.

## References


1. Najjar, j., Klobučar, T., Nguyen-Ngoc,A. V., Totschnig, M. , Müller, F., Simon, B., Karlsson, M., Eriksson, H. Towards Outcome Based Learning: An Engineering Education Case.
2. Salles, T., Rocha, L., Pappa, G., Mourao, G., Meira , W. Jr. , Goncalves, M.( 2010) Temporally-aware algorithms for document classification. ACM SIGIR Conference.
3. Najjar,J. (Editor), Simon, B. (WUW),Klobučar, T. (JSI), ISURE: Report on implementation and validation of a standardized model for learning needs analysis and the provision of learning opportunities. (ICOPER).
4. Najjar,J. (WUW), Grant,S. (JISC CETIS), Simon,B. (WUW), Derntl, M. (UNIVIE), Klobu"ar,T. (JSI),Crespo, M, Kloos, C. K. (UC3M), Nguyen-Ngoc, A.V. (ULE), Pawlowski, J. (JYV) and Hoel, T. (OUC), Oberhuemer, P. (UNIVIE). ISURE: Model for describing learning needs and learning opportunities taking context ontology modeling into account.
5. Najjar, J., Ternier, S., Erik.Duval.Learning Object Metadata: Opportunities and Challenges for the Middle East and North Africa.
6. Osmar, R. Za¨ıane University of Alberta, Edmonton, Alberta, Canada.Building a Recommender Agent for e-Learning Systems.
7. Frankowski, D., Shyong, K. (Tony) Lam,Sen, S. , F. Maxwell Harper, Yilek, S., Cassano, M.,Riedl, J., Recommenders Everywhere: The WikiLens Community-Maintained Recommender System.
8. S. A. Macskassy, F. Provost. Classification in Networked Data: A Toolkit and a Univariate Case Study, Journal of Machine Learning Research, Vol. 8, pp. 935–983, 2007.
9. Kale, S., Fengyun, E. H., Cao Jaswinder Pal Singh. Analysis and Algorithms for Content-based Event Matching.
10. Imran, K., Ghauth Multimedia University,Nor Aniza Abdullah University of Malaya. Measuring learner's performance in e-learning recommender systems.
11. Imran, K., Ghauth. B, Abdullah, N. A., Building an E-Learning Recommender System using Vector Space Model and Good Learners Average Rating.
12. Michael J. Pazzani1 and Billsus, D., Content-Based Recommendation Systems.
13. Najjar,J. , Klobučar, T. , Nguyen-Ngoc, A. V.,Totschnig M. , Müller, F. , Simon, B. , Karlsson, M., Eriksson, H., Towards Outcome Based Learning: An Engineering Education Case.
14. H. Raghavan, J. Allan. An interactive algorithm for asking and incorporating feature feedback into support vector machines. ACM SIGIR Conference, 2007.
15. E. S. Ristad, and P. N. Yianilos, Learning String Edit Distance. Tech. Rep. CS-TR-532-96, Department of Computer Science, Princeton University, Princeton, N.J.October 1996. Revised November 1997.
16. Imran, K., Ghauth Multimedia University,Nor Aniza Abdullah University of Malaya. Measuring learner's performance in e-learning recommender systems.
17. Imran, K. , Ghauth, B. , Abdullah, N. A., Building an E-Learning Recommender System using Vector Space Model and Good Learners Average Rating.
18. Y.-H. Kim, S.-Y. Hahn, B.-T. Zhang. Text filtering by boosting naive Bayes classifiers. ACM SIGIR Conference, 2000.
19. W. Lam, K.-Y. Lai. A meta-learning approach for text categorization. ACM SIGIR Conference, 2001.

20. D. Mladenic, J. Brank, M. Grobelnik, N. Milic-Frayling. Feature selection using linear classifier weights: interaction with classification models. ACM SIGIR Conference, 2004. [100] K. Myers, M. Kearns, S. Singh, M. Walker. A boosting approach to topic spotting on subdialogues. ICML Conference, 2000.