



國立中央大學  
*National Central University*

## INTERNATIONAL SYMPOSIUM ON GRIDS & CLOUDS 2018

# SKILL-BASED OCCUPATION RECOMMENDATION

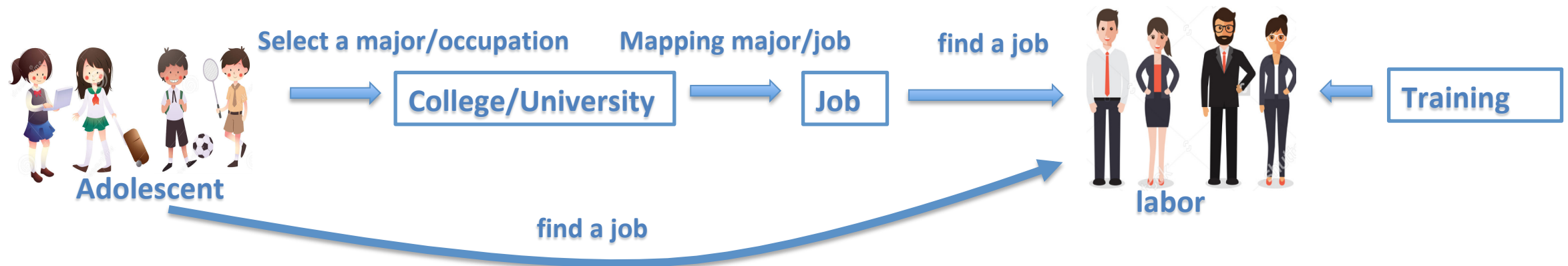
Ankhtuya Ochirbat,  
Timothy K. Shih,

National University of Mongolia, Mongolia  
National Central University, Taiwan

Presenter: O. Ankhtuya  
2018.03.21

# Introduction

- A major choice in **high school** or **undergraduate stage** is an important decision in the person life.
- When students choose the college major **generally**, they first intend and select the **occupation** that they will work through it in the future.
- But the some **occupations** are not clear **to map into the academic program** to study or vice versa.



# Recommender Systems

Recommendation system is an information filtering technique, which provides users with information, which he/she may be interested in.

## Examples:



Video-on-demand provider in North America and UK

- Matches 23 million customers with a huge inventory of movies according to their tastes
- 60 -70% of views result from the recommendations<sup>9</sup>



Gold standard of e-commerce. Pioneer in using recommendations

- Sits on a huge volume of collective information of its customers
- Customers can view what people with similar tastes viewed or purchased
- Customers can ask the recommendations engine to ignore selected purchases



Social and professional networking sites

- Sits on a huge volume of collective information of its customers
- Customers can view what people with similar tastes viewed or purchased
- Customers can ask the recommendations engine to ignore selected purchases

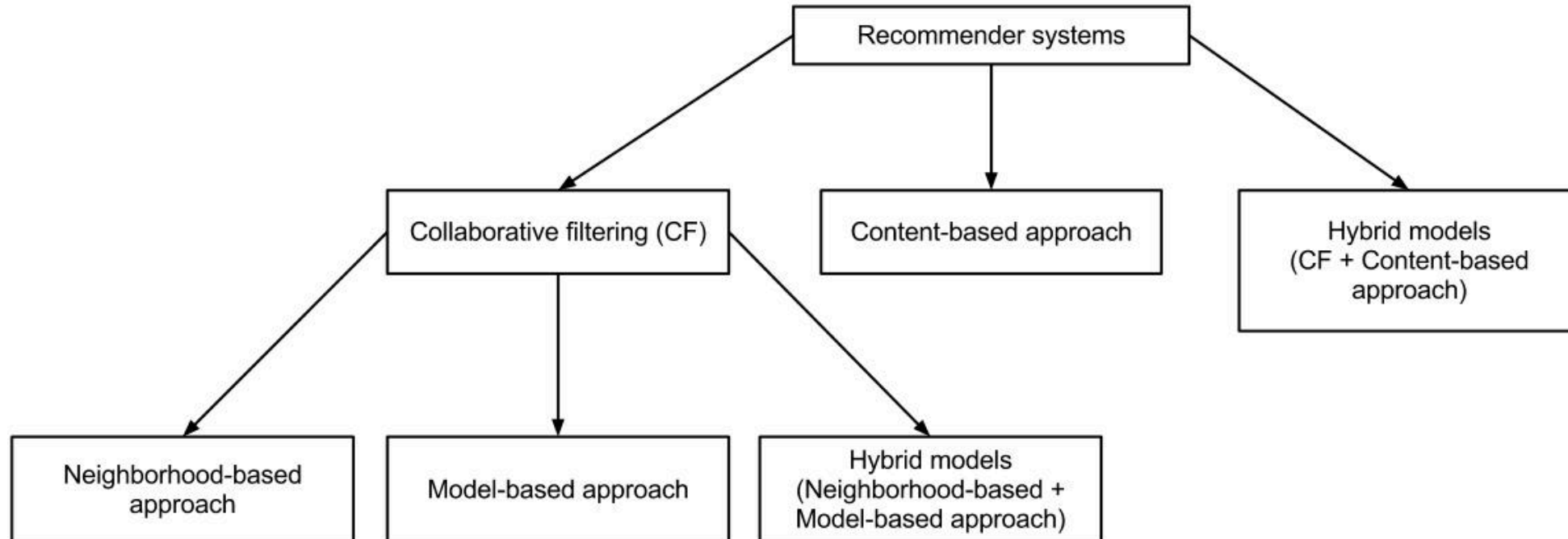


Music station. Offers music suggestions based on ratings

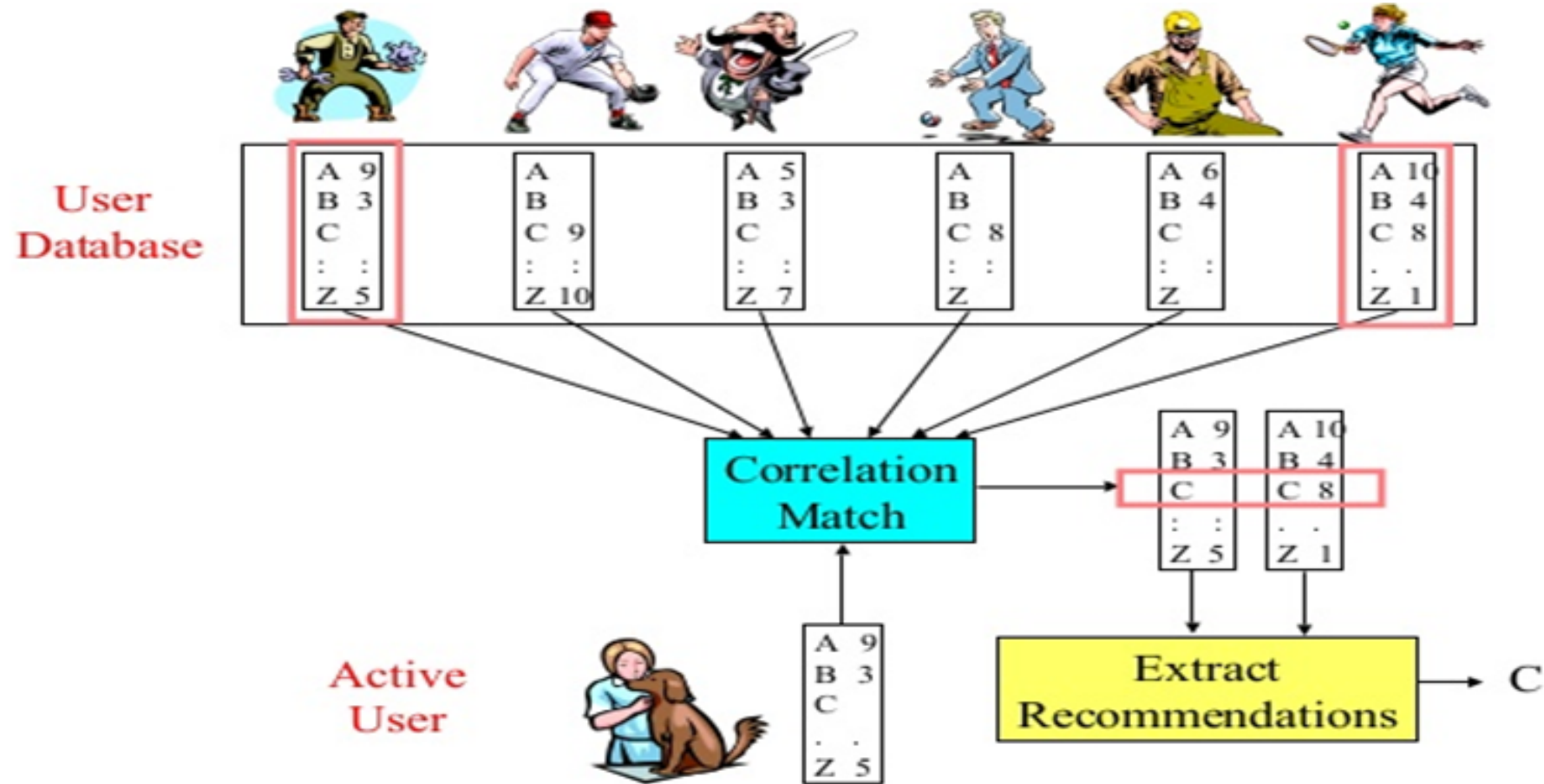
- Sits on a huge volume of collective information of its customers
- Customers can view what people with similar tastes viewed or purchased
- Customers can ask the recommendations engine to ignore selected subscriptions<sup>3</sup>

# Recommender Systems

- Recommender Systems can be broadly categorized as



# Collaborative Filtering

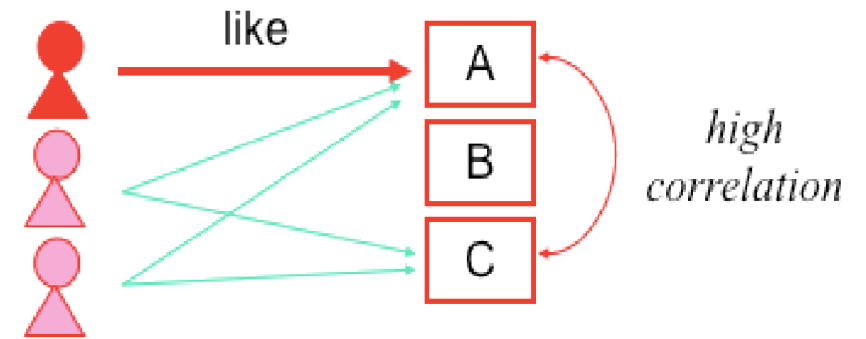


- Basic Idea- Recommend items that are similar to the user's highly preferred items.

# Collaborative Filtering

- **Item Based Collaborative Filtering**

- Use user-item ratings matrix
- Make item-to-item correlations
- Find items that are highly correlated
- Recommend items with highest correlation



- **Similarity Metric :**

$$itemSim(i, j) = \frac{\sum_{u \in RB_{i,j}} (r_{ui} - \bar{r}_u)(r_{uj} - \bar{r}_u)}{\sqrt{\sum_{u \in RB_{i,j}} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{u \in RB_{i,j}} (r_{uj} - \bar{r}_u)^2}}$$

- **Prediction Function :**

$$pred(u, i) = \frac{\sum_{j \in ratedItems(u)} itemSim(i, j) \cdot r_{uj}}{\sum_{j \in ratedItems(u)} itemSim(i, j)}$$

# Content-based Approach

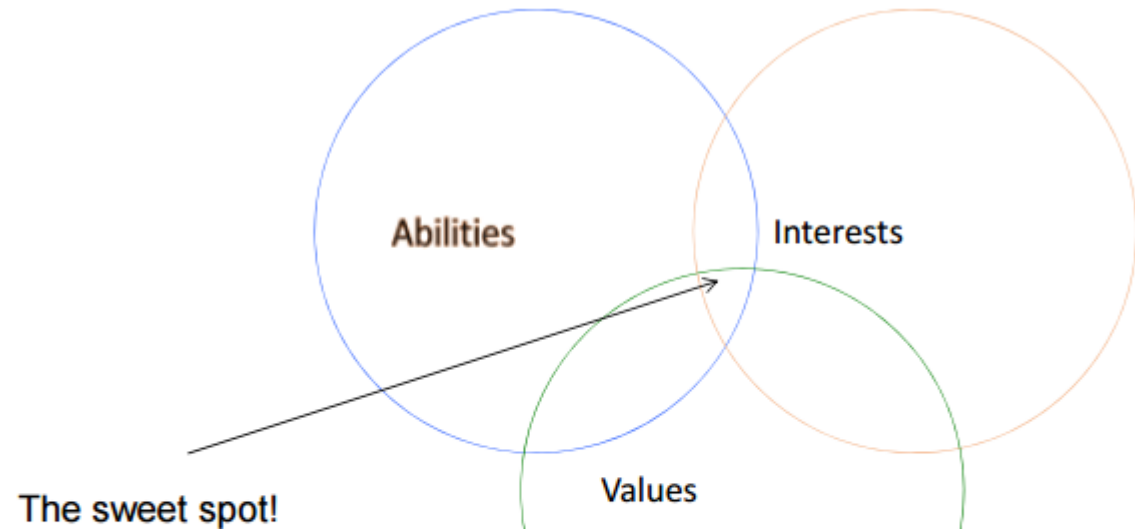
- These approaches recommend items that are similar in content to items the user has liked in the past, or matched to attributes of the user.
- **Proposed method: TF/IDF** - Term Frequency / Inverse Document Frequency
  - Term Frequency - frequency of occurrence of a term in a given document.
  - Inverse Document Frequency - measure of the general importance of the term.

$$TF_{i,j} = \frac{f_{i,j}}{\max_z f_{z,j}} \quad IDF_i = \log \frac{N}{n_i} \quad w_{i,j} = TF_{i,j} \times IDF_i$$

Where, the maximum is computed over the **frequencies**  $f_{z,j}$  of all **keywords**  $k_z$  that appear in **occupation detail**  $d_j$ . The measure of **inverse document frequency** ( $IDF_i$ ) is applied in combination with simple **term frequency** ( $TF_{i,j}$ ).

# Related works

- Gordon's curriculum for working with undecided students
  - Self-assessment
  - Educational planning
  - Career planning
  - Decision-making



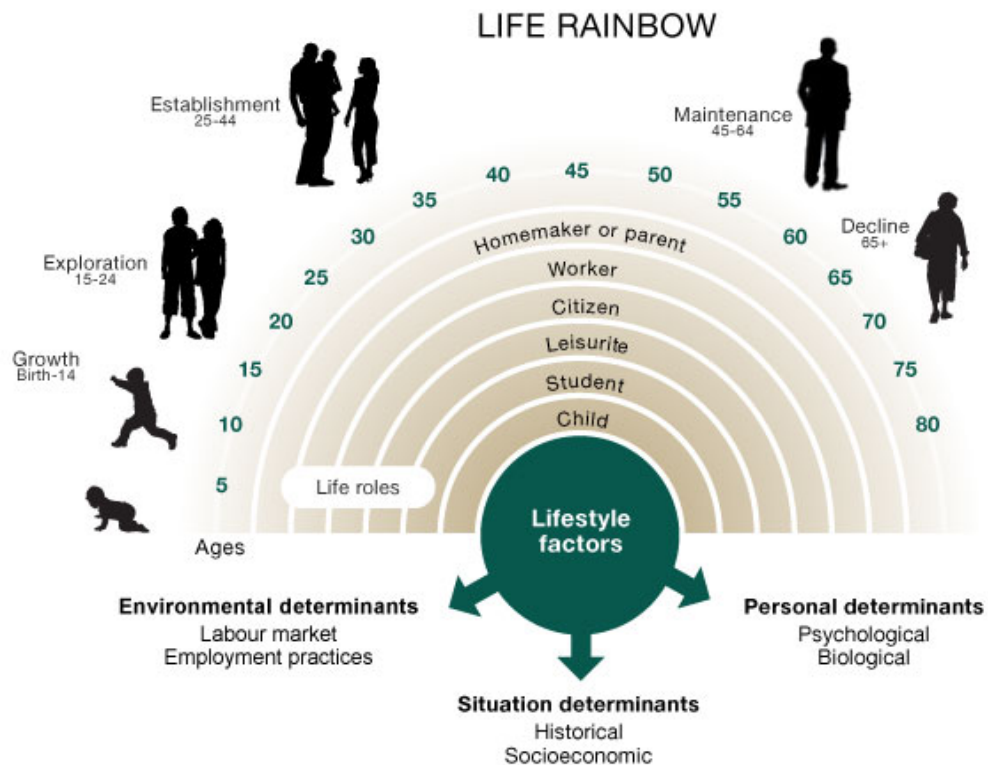
## The Relationships

- **Interests:** Things we like to do
- **Abilities/Skills:** Things we do well
- **Values:** That which is most important to us



# Related works

- Career counseling theory and adolescents
  - Super's theory of career development



Super's five life and career development stages

Stage	Age	Characteristics
Growth	birth-14	Development of self-concept, attitudes, needs and general world of work
Exploration	15-24	"Trying out" through classes, work hobbies. Tentative choice and skill development
Establishment	25-44	Entry-level skill building and stabilisation through work experience
Maintenance	45-64	Continual adjustment process to improve position
Decline	65+	Reduced output, prepare for retirement

# Background and Related works

---

- We built two kinds of **career** and **occupation recommendation systems** and **conducted** the **experiments** among the **high school** and the **college students** of Mongolia, Taiwan and other countries.
  - **Career Recommendation System** in 2014/2015 academic year
  - **Occupation Recommendation System** in 2015/2016 academic year

# Background and Related works

---

- **Methods for both systems**

- **Hybrid Recommendation** techniques were employed. Namely:
  - Collaborative Filtering (CF):
  - Content-based Filtering
  
- **There are 3 main steps:**
  1. Data were normalized.
  2. Similarities were computed.
  3. Predictions/Recommendations were calculated.

# Problem Statement

- In order to help students in major choice, it is essential to build the occupation recommendation system for the student with a capacity to meet all the needs
  - where it provide direction and guidance to students in choosing a major that suits with their interests, skills and abilities.



are interested in



**engineering occupation**

**What is engineer?**

**What do they do?**

**What kind of engineers? ...**

# Classification of Instructional Programs

---

- Classification of Instructional Programs (CIP) is a taxonomic coding scheme of instructional programs
  - › Morgan, R. L. (1991). Classification of instructional programs
  - › CIP Canada 2016

## Detail for CIP Code 11.0102



**Title:** Artificial Intelligence.

**Definition:** A program that focuses on the symbolic inference, representation, and simulation by computers and software of human learning and reasoning processes and capabilities, and the computer modeling of human motor control and motion. Includes instruction in computing theory, cybernetics, human factors, natural language processing, and applicable aspects of engineering, technology, and specific end-use applications.

See also: [14.4201\) Mechatronics, Robotics, and Automation Engineering.](#)

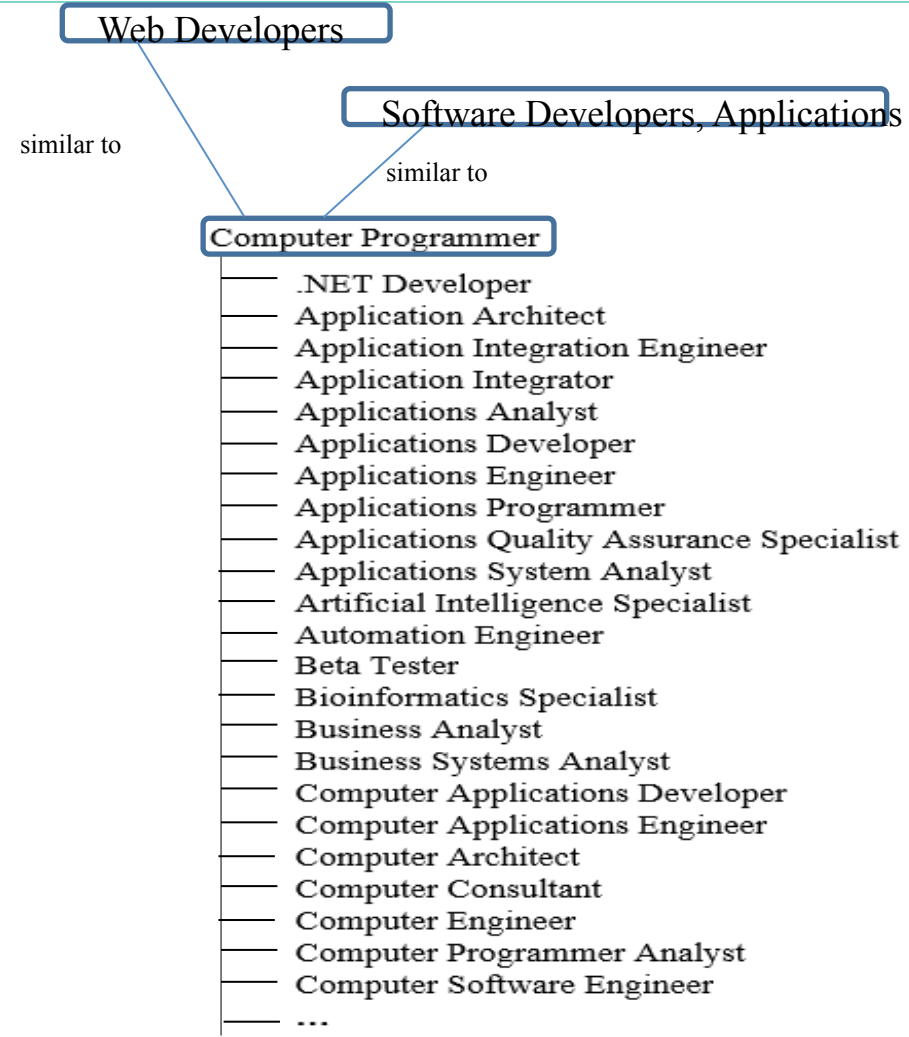
# Job zone

- Job Zones group occupations
  - levels of education, experience, and training necessary to perform the occupation.

**There are five job zones:**

Job Zone	Preparation Needed	Example	Degrees/Experience
5	Extensive	Doctors	Doctorate's and master's with experience
4	Considerable	Teachers	Bachelor's and master's
3	Medium	Electricians	Associate's, bachelor's, and apprenticeships
2	Some	Tellers	some on-the-job training
1	Little to no	Waiters	minimal on-the-job training

# Semantic search



An excerpt of dictionary of occupational titles in Computer programmer.

doctor 

- Bioinformatics Scientists
- Medical Scientists, Except Epidemiologists
- Chiropractors
- Dentists, General
- Prosthodontists
- Anesthesiologists
- Family and General Practitioners
- Internists, General
- Pediatricians, General
- Allergists and Immunologists
- Dermatologists
- Nuclear Medicine Physicians
- Ophthalmologists
- Physical Medicine and Rehabilitation Physicians
- Sports Medicine Physicians
- Urologists
- Podiatrists
- Audiologists
- Naturopathic Physicians
- Medical Assistants

Example of semantic search. An adolescent searched a doctor as an occupation. In below of it occupations are listed which is related the keyword semantically.

# Preprocessing

- HTML

## Lawyer

From Wikipedia, the free encyclopedia

A **lawyer** is a person who practices law, as an advocate, barrister, attorney, counselor or solicitor or chartered legal executive.<sup>[1]</sup> Working as a lawyer involves the practical application of abstract legal theories and knowledge to solve specific individualized problems, or to advance the interests of those who hire lawyers to perform legal services.

The role of the lawyer varies greatly across legal jurisdictions, and so it can be treated here in only the most general terms.<sup>[2][3]</sup>

- Text n

- Tokenization

- Remo

- Stem

`<p>A <b>lawyer</b> is a person who <a href="/wiki/Practice_of_law" title="Practice of law">practices law</a>, as an <a href="/wiki/Advocate" title="Advocate">advocate</a>, <a href="/wiki/Barrister" title="Barrister">barrister</a>, <a href="/wiki/Attorney_at_law" title="Attorney at law">attorney</a>, <a href="/wiki/Counsel" title="Counsel">counselor</a> or <a href="/wiki/Solicitor" title="Solicitor">solicitor</a> or <a href="/wiki/Legal_executive" title="Legal executive">chartered legal executive</a>.<sup id="cite_ref-1" class="reference"><a href="#cite_note-1">[1]</a></sup> Working as a lawyer involves the practical application of abstract legal theories and knowledge to solve specific individualized problems, or to advance the interests of those who hire lawyers to perform legal services.</p>  
<p>The role of the lawyer varies greatly across legal jurisdictions, and so it can be treated here in only the most general terms.<sup id="cite_ref-2" class="reference"><a href="#cite_note-2">[2]</a></sup><sup id="cite_ref-3" class="reference"><a href="#cite_note-3">[3]</a></sup></p>`

## Lawyer

A lawyer is a person who practices law, as an advocate, barrister, attorney, counselor or solicitor or chartered legal executive. Working as a lawyer involves the practical application of abstract legal theories and knowledge to solve specific individualized problems, or to advance the interests of those who hire lawyers to perform legal services. The role of the lawyer varies greatly across legal jurisdictions, and so it can be treated here in only the most general terms.



# Preprocessing

- HTML parser
- Text mining - preprocessing
  - Tokenization
  - Removing stop words
  - Stemming

A lawyer is a person who practices law, as an advocate, barrister, attorney, counselor or solicitor or chartered legal executive.



A lawyer is a person who practices law , as  
an advocate , barrister , attorney , counselor  
or solicitor or chartered legal executive

# Preprocessing

- HTML parser
- Text mining
  - Tokenization
  - Removing stop words
    - Regular expression
  - Stemming

A lawyer is a person who practices law , as  
an advocate , barrister , attorney , counselor  
or solicitor or chartered legal executive

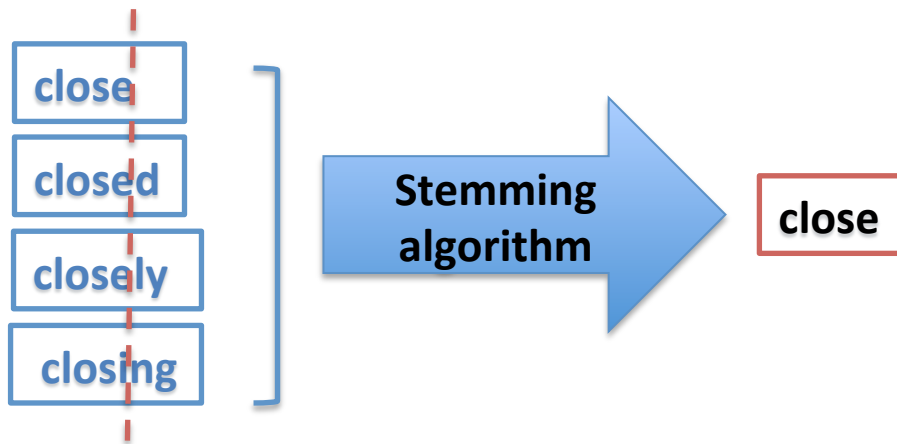
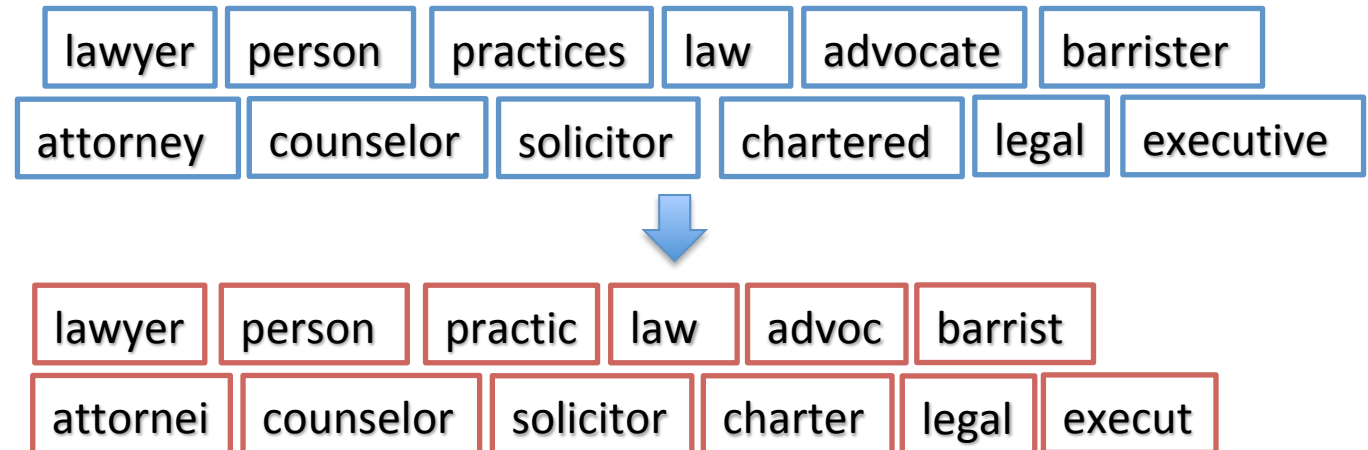


lawyer person practices law advocate barrister  
attorney counselor solicitor chartered legal executive

```
$stopwords = array("a", "about", "above", "above", "across", "after", "afterwards", "again", "against", "all", "almost", "alone", "along", "already", "also", "although", "always", "am", "among", "amongst", "amount", "amountst", "amounts", "an", "and", "another", "any", "anyhow", "anyone", "anything", "anyway", "anywhere", "are", "around", "as", "at", "back", "be", "became", "because", "become", "becomes", "becoming", "been", "before", "beforehand", "behind", "being", "below", "beside", "besides", "between", "beyond", "bill", "both", "bottom", "but", "by", "call", "can", "cannot", "cant", "co", "con", "could", "couldnt", "cry", "current", "cv", "day", "de", "describe", "detail", "do", "done", "down", "due", "during", "each", "eg", "eight", "either", "eleven", "else", "elsewhere", "empty", "enough", "etc", "even", "ever", "every", "everyone", "everything", "everywhere", "except", "excit", "few", "fifteen", "fify", "fill", "find", "fire", "first", "five", "for", "former", "formerly", "forty", "found", "four", "from", "front", "full", "further", "junior", "get", "give", "go", "had", "has", "hasnt", "have", "he", "hence", "her", "here", "hereafter", "hereby", "herein", "hereupon", "hers", "herself", "him", "himself", "his", "how", "however", "hundred", "ie", "if", "in", "inc", "indeed", "interest", "into", "is", "it", "its", "itself", "keep", "kindly", "last", "latter", "latterly", "least", "like", "less", "ltd", "made", "make", "many", "may", "me", "meanwhile", "might", "mill", "mine", "more", "moreover", "most", "mostly", "move", "much", "must", "my", "myself", "name", "namely", "neither", "never", "nevertheless", "next", "new", "nine", "no", "nobody", "none", "noone", "nor", "not", "nothing", "now", "nowhere", "occupation", "occupations", "of", "off", "often", "on", "once", "one", "only", "onto", "or", "other", "others", "otherwise", "our", "ours", "ourselves", "out", "over", "own", "part", "per", "perhaps", "please", "put", "rather", "re", "same", "see", "seem", "seemed", "seeming", "seems", "send", "senior", "serious", "several", "she", "should", "show", "side", "since", "sincere", "six", "sixty", "skill", "so", "some", "somehow", "someone", "something", "sometime", "sometimes", "somewhere", "still", "such", "system", "take", "ten", "than", "that", "the", "their", "them", "themselves", "then", "thence", "there", "thereafter", "thereby", "therefore", "therein", "thereupon", "these", "they", "thickv", "thin", "third", "this", "those", "though", "three", "through", "throughout", "thru", "thus", "to", "together", "too", "top", "toward", "towards", "twelve", "twenty", "two", "un", "uk", "under", "until", "up", "upon", "us", "very", "via", "was", "we", "well", "were", "what", "whatever", "when", "whence", "whenever", "where", "whereafter", "whereas", "whereby", "wherein", "whereupon", "wherever", "whether", "which", "while", "whither", "who", "whoever", "whole", "whom", "whose", "why", "will", "with", "within", "without", "would", "year", "yet", "you", "your", "yours", "yourself", "yourselves", "the", 'www!');
```

# Preprocessing

- HTML parser
- Text mining
  - Tokenization
  - Removing stop words
  - Stemming

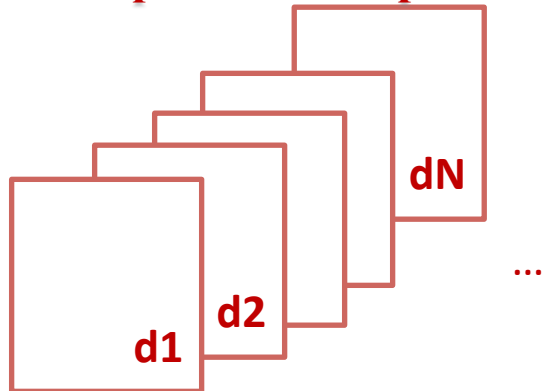


# Analysis - Text mining

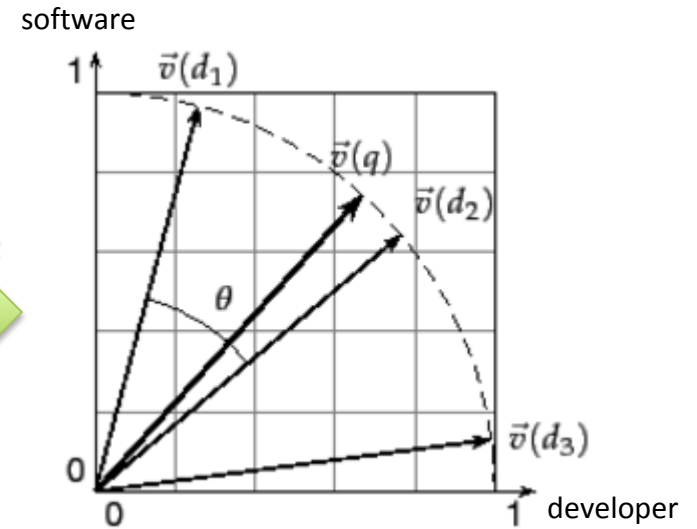


Query: Software developer

occupation descriptions



Preprocessing & TF-IDF



$$\text{sim}(d_1, d_2) = \frac{\vec{V}(d_1) \cdot \vec{V}(d_2)}{|\vec{V}(d_1)| |\vec{V}(d_2)|}$$

# Result

## *Student's intended occupation and its relevant wiki occupation*

No.	Intended Occupation	Relevant Wiki occupation title	Relevant Wiki categories
1	business manager	General manager	Management occupations
2	economist	Chief economist	Business occupations
3	fitness teacher	Substitute teacher	Education and training occupations
4	designer	Costume designer	Fashion occupations
5	engineer	Systems engineering	Engineering occupations
6	doctor, engineer	Systems engineering	Engineering occupations
7	lawyer	Cause lawyer	Legal professions
8	doctor, lawyer	Cause lawyer	Legal professions
9	athlete	Sports agent	Business occupations
10	practitioner engineer	Systems engineering	Engineering occupations
11	civil enigeer	First Civil Service Commissioner	Government occupations
12	police officer	Law enforcement officer	Legal professions
13	veterinarian	Zoological medicine	Healthcare occupations
14	captain	Captain	Occupations
15	marine captain, manager	Captain	Occupations

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn} = 0.93$$

Where true positive ( $tp$ ), true negative ( $tn$ ), false positive ( $fp$ ), and false negative ( $fn$ )

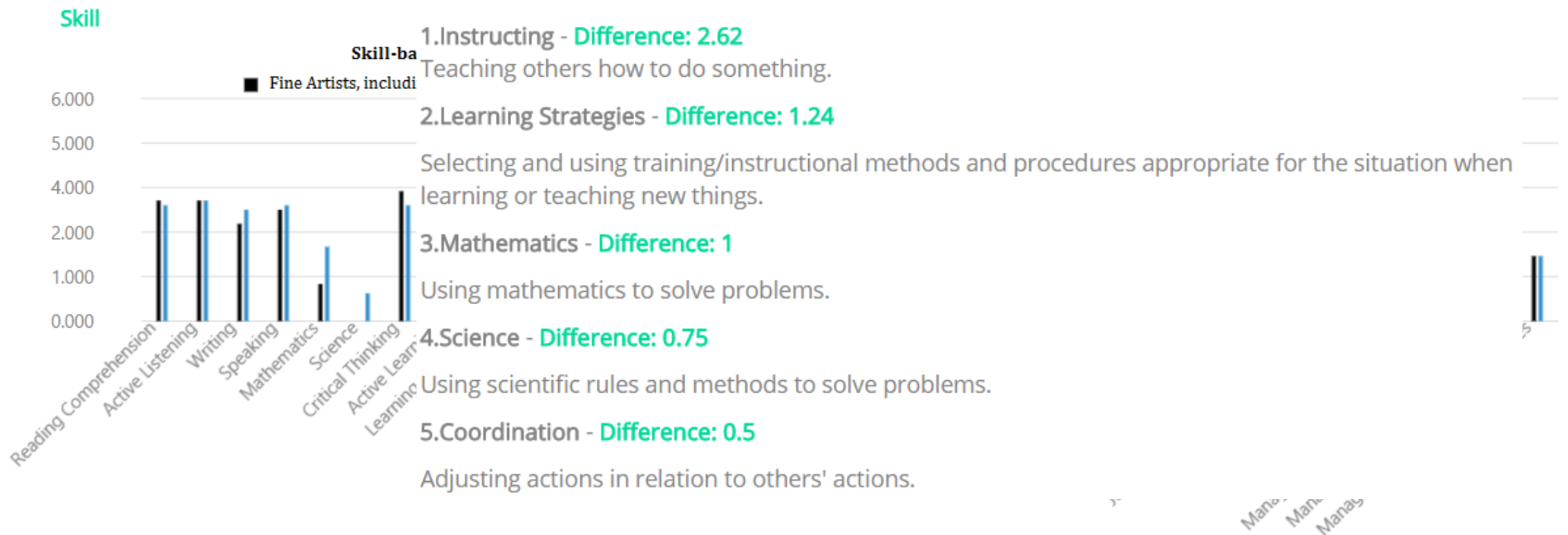
# Result

- **Occupation relatedness**

Dataset	Accuracy	Precision	Recall	F1
1.Occupations (O*NET)	.94	.96	.94	.95
2.Wiki occupations	.95	.98	.96	.97
3.Training set – Occupations Test set – Wiki occupation	.90	.91	.91	.90
4.Training set – Wiki occupation Test set – Occupations	.61	.66	.60	.54

# Skill gaps

- A skill gap method is employed to compare the differences in skills between the intended-based occupation and an occupation from the skill questionnaire since our users are adolescents without any job experiences



# Skill gaps

- Mean Absolute Error (MAE) measures an average magnitude of the errors without considering their direction.

MAE of skills between skill-based and intended-based occupations for all students

Skill name	MAE	Skill name	MAE	Skill name	MAE
1.Equipment Selection	0.91	13.Systems Evaluation	0.74	25.Negotiation	0.55
2.Troubleshooting	0.9	14.Systems Analysis	0.71	26.Technology Design	0.55
3.Mathematics	0.88	15.Management of Financial Resources	0.7	27.Persuasion	0.55
4.Science	0.88	16.Service Orientation	0.67	28.Active Listening	0.53
5.Quality Control Analysis	0.85	17.Speaking	0.66	29.Programming	0.5
6.Operations Analysis	0.84	18.Management of Personnel Resources	0.64	30.Critical Thinking	0.5
7.Instructing	0.83	19.Management of Material Resources	0.63	31.Social Perceptiveness	0.49
8.Operation and Control	0.83	20.Judgment and Decision Making	0.6	32.Coordination	0.48
9.Equipment Maintenance	0.81	21.Complex Problem Solving	0.58	33.Installation	0.42
10.Repairing	0.8	22.Active Learning	0.58	34.Monitoring	0.41
11.Operation Monitoring	0.79	23.Writing	0.58	35.Time Management	0.36
12.Learning Strategies	0.76	24.Reading Comprehension	0.57		



# Usability

- To validate usefulness of online course with ORS, System Usability Scale (SUS) was employed.
  - 10 items with responses made on a Likert scale format ranging from 1 = strongly disagree to 5 = strongly agree

$$\text{Average degree} = \sum_{m=23}^{j=1} \left( \sum_{n=10}^{i=1} \begin{cases} \text{score}_{j,i} - 1, & \text{if } i \% 2 == 1 \\ 5 - \text{score}_{j,i}, & \text{else} \end{cases} * 2.5 \right) / m$$

Where,  $\text{score}_{j,i}$  is the rating of student  $j$  on item  $i$ . And,  $n$  is the number of questions.

# Usability

System Usability Scale		
No	Item	Average
1	I think that I would like to use this system frequently	3.75
2	I found the system unnecessarily complex	3.5
3	I thought the system was easy to use	4.5
4	I think that I would need the support of a technical person to be able to use this system	2
5	I found the various functions in this system were well integrated	4
6	I thought there was too much inconsistency in this system	1.88
7	I would imagine that most people would learn to use this system very quickly	4.13
8	I found the system very cumbersome to use	1.56
9	I felt very confident using the system	3.69
10	I needed to learn a lot of things before I could get going with this system	2.13
	Usability – Average degree	72.5%

# Conclusion

---

- The aim of this study was to implement Skill-based Occupation Recommendation Systems (SORS) and to apply it in an effort to improve major/career plans of adolescents.
  - Skill-gap
  - Usability of SORS
- In the future, we will conduct an online course in the career counselling session using MOOC and Wiki Education Foundation, and to track students' interested learning directions through variety of subjects.
- Another future study is to build an iterative dialogue system according to this proposed system's improvement.
  - The system can popup interactive dialogs with the student, and to ask additional questions after providing recommendations.



Thank you for your attention.

[ankhaa8@gmail.com](mailto:ankhaa8@gmail.com)