International Journal of Computer Applications Technology and Research
Volume 7–Issue 06, 203-207, 2018, ISSN:-2319–8656

An Empirical Study on the Relationship between Economic Openness and Economic Growth in China

Mengshan Zhu
China Agriculture University
Beijing, China

Abstract: Economic openness is the measure of economic activity in the country comprehensive index. How is economic openness indicator measured? Chinese economy has experienced rapid growth for more many years, what is on earth the effect of economic opening on Chinese economic growth? The answer to this question will provide instructive revelation about the selection of Chinese reasonable opening policy. Economic openness is measured by trade openness, foreign investment openness and financial openness in this paper. Based on Solow economic growth model and beginning with foreign trade, foreign investment and financial development, this paper made regression analysis using Chinese data from 1985 to 2004. The empirical analysis indicates that the domestic capital input is still the primary element that promotes Chinese economic growth, by contrast, the effect of foreign trade and foreign investment on Chinese economic growth is faint. Again, financial development on the impetus of economic growth in China has a room to rise.

Keywords: Economic opening; Financial development; Chinese economic growth; Empirical analysis

1. INTRODUCTION

Since the Chinese Economic Reform and open up, China's economic development has made remarkable achievements. Foreign trade and foreign investment have increased significantly. The level of financial development has also been continuously improved. The degree of integration of finance and trade in the real economy has also deepened, and future economic growth will depend on more effective expand and deepen opening to the outside world. However, opening up to the outside world is a double-edged sword. On the one hand, it will bring economic benefits to a country and it will also need to pay a corresponding price. The process of economic opening in each country is the coexistence of benefits and costs, besides opportunities and challenges. Therefore, the problem we face is not whether to open to the outside world, but how to effectively and rationally implement opening up to the outside world. Specifically, we should evaluate whether a country's openness is based on whether it is conducive to promoting domestic economic growth. However, all along, new methods and viewpoints have emerged in the field of measurement and measurement of openness, and the practice and understanding of openness have continued to enter a new stage. The requirements of foreign economic opening have involved exchange rate policy, monetary policy, the finance system and macroeconomic. Therefore, this paper constructs a new openness index system from three aspects of trade openness, foreign capital openness and financial openness, and adopts data from 1985 to 2004 on economic growth and the empirical analysis was conducted on the relationship between openness to the outside world.

2. A REVIEW OF THE RESEARCH ON THE RELATIONSHIP BETWEEN OPENING UP AND ECONOMIC GROWTH

The research on the relationship between the degree of opening up to the outside world and economic growth has a long history. In general, most of the viewpoints believe that opening to the outside world can promote economic growth. However, there are significant differences in the views of different schools on the specific process of opening up to economic growth and the internal mechanism.

2.1 Theoretical Research Review

Earlier economists looked at the relationship between openness and economic growth mainly from the perspective of foreign trade. In the 1930s, British scholar Robertson (D.H. Robertson) proposed for the first time that foreign trade is the engine for economic growth (engine for growth) proposition. The main focus is that the backward countries can promote their own economic growth through foreign trade, especially export growth. Harrison analyzed the relationship between trade openness and economic growth using seven indicators of trade openness. [1] In 2003, domestic scholars Bao Qun et al. selected the five indicators of trade dependence, actual tariff rate, black market transaction fees, Daulas index and revised trade dependence to measure the degree of trade openness to economic growth since China’s reform and opening up. [2] Foreign direct investment (FDI) mainly affects its economic growth by promoting the capital accumulation and technological progress of the host country. The conclusion that FDI contributes to the accumulation of capital in the host country stems from the double-gap theory proposed by Chen (1966). DeMello (1999) studied the relationship between FDI and economic growth in OECD countries. The conclusion is that only when FDI and domestic investment are complementary, FDI has a positive impact on economic growth. He Zhengxia started from the perspectives of foreign investment and foreign trade, and examined the actual effect of economic opening on China’s economic growth. [3]

2.2 Empirical Research Overview

Early empirical research on the relationship between external development and economic growth mainly focused on the test of export-led economic growth (ELG). In the past 20 years, scholars have conducted extensive research using different models and analysis methods for different countries and regions. The relevant studies can be divided into the following two categories.

(1) This type of research uses cross-country (regional) cross-sectional data and uses rank correlation tests and OLS regression methods to directly analyze the relationship between external and GDP, or join labor, capital, and investment with such elements as export factors, the OLS regression method is used to analyze the impact of exports on economic growth. The number of selected countries (regions) is not equal, and different periods are used. For example, Luo Zhongzhou used panel data from 1994 to 2005 to conduct a comparative study of the relationship between openness and economic growth in the Pearl River Delta, the Yangtze River Delta, and the Bohai Rim. [4]

(2) In recent years, time series analysis based on individual country data has become mainstream, and the most widely used analysis method is the Granger causality test. Some of the variable systems considered in this type of research only examined the bivariate systems of openness and GDP, and some added variables such as investment, technological progress, and exchange rate in the bivariate system. Because of the differences in model methods and the selection of variables, the conclusions of such empirical research are inconsistent. For example, Lan Yisheng empirically analyzed the relationship between the degree of openness and economic growth in various regions of China from 1985 to 1998 and believed that opening to the outside world has strongly promoted the growth of China's national economy. [5]

3. CONSTRUCTION AND ANALYSIS OF OPENNESS INDEX SYSTEM

The degree of economic openness is an important indicator for measuring the level of open economy. At present, the research on the relationship between openness and economic growth has received increasing attention from the academic community, but no consensus has been reached on the selection of indicators for economic openness. Earlier economists, when measuring openness to the outside world, often only examined the dependence on foreign trade, that is, the ratio of total import and export trade to gross domestic product (GDP). Although this method is intuitive and simple, in the course of the study, people gradually measure the degree of openness with the degree of trade dependence, because a country’s trade dependence is affected by factors such as the country’s capital level, financial conditions, and government policies. Therefore, the degree of trade dependence does not fully reflect the level of openness. Since then, more scholars have measured the degree of openness from the perspective of international openness in foreign trade, foreign direct investment, etc., and studied the internal relationship with economic growth, thus making great progress. However, with the acceleration of the free flow of financial capital in the world, virtual capital plays an increasingly important role in the foreign economy. Therefore, based on previous research, this paper integrates financial openness into the economic openness index system. The formula is as follows:

China's economic openness = (trade openness + foreign capital openness + financial openness) /3

which is

$$\mathrm{EO} = (TO + IO + FO)/3 \qquad (1)$$

Among them, EO is China's economic openness, TO is China's trade openness, IO is China's foreign capital openness, and FO is China's financial openness.

$$\mathrm{TO} = TO_{Goods} + IO_{services} = \frac{Commodity\ trade\ import\ and\ export + \text{Total import and export of service trade}}{GDP} \qquad (2)$$

$$\mathrm{IO} = IO_{accept} + IO_{outside} = \frac{China\ actually\ accepts\ total\ foreign\ direct\ investment + \text{Total foreign direct investmen}}{GDP} \qquad (3)$$

$$\mathrm{FO} = (FO_{currency} + FO_{capital})/2$$

$$FO_{capital} = FO_{Securities} + FO_{other}$$

which is

$$\mathrm{FO} = \frac{(FO_{currency} + FO_{Securities} + FO_{other})}{2} = \left(\frac{Central\ Bank' s\ foreign\ net\ assets}{Central\ Bank's\ total\ assets} + Total\ securities\ investment + Total\ other\ investment GDP/2 \qquad (4)$$

The above indicators comprehensively reflect the level of a country’s economic internationalization from the perspective of breadth [6]. The TO indicator is actually an opening degree of a state-owned commodity market and an intangible commodity market. The openness of the intangible commodity market is often overlooked by people. However, the importance of this market to the national economy has been greatly improved and can’t be ignored. IO reflects the openness of China's foreign capital market. It should be noted that in order to highlight the importance of FDI, the IO in this article refers to the opening of direct investment, and does not include other investment in securities investment. Our country accepts the openness of foreign investment and our country directly The openness of investment constitutes. In addition, securities investments and other investments are an important part of the economic activities in the capital market and cannot be ignored. This article is included in the open financial sector. The FO reflects the opening level of China's financial market, which consists of two indicators: the openness of the currency market and the openness of the capital market. Among them, the degree of capital openness is again composed of two sub-indicators: the degree of openness of securities investment and the degree of other investment. The FO currency represents the opening level of a country's currency market. Here, the ratio of foreign assets of the central bank to its total assets is used as a measure to measure the importance of foreign currency in the domestic currency supply.

According to the statistical yearbook data of the relevant year, China's trade openness, foreign capital openness, financial openness, and openness were calculated from 1985 to 2004. It can be seen that since the reform and opening up, the overall level of China’s openness has been continuously rising.

Table 1 Related data of openness (Unit: Billion US Dollars)

Year	GDP	Commodity trade import and export	Total import and export of service trade	Absorb direct investment	Foreign direct investment	Central Bank's total assets (billion yuan)	Central Bank's foreign net assets (billion yuan)	Total securities investment	Total Other investment
1985	3052.6	696.0	51.9	19.6	6.3	273.23	14.57	20.27	18.56
198	2954.8	606.3	56.4	18.8	4.5	310.77	15.48	16.48	36.07
198	3213.9	711.3	65.2	23.1	6.5	372.73	21.17	13.31	32.81
1988	4010.7	874.2	80.2	31.9	8.5	451.46	24.45	15.56	60.93
198	4491.0	1116.8	79.7	33.9	7.8	563.80	36.21	4.6	17.48
1990	3877.7	1154.4	98.1	34.9	8.3	830.47	79.60	2.41	61.98
1991	4061.0	1357.0	107.3	43.5	9.1	777.45	169.96	8.95	46.5
1992	4830.5	1655.3	182.4	108.9	40.0	935.16	133.04	8.43	73.49
1993	4018.5	1957.0	224.9	274.7	44.0	1266.80	145.99	42.44	26.90
1994	5425.3	2366.2	319.2	337.7	20.0	1758.80	445.13	43.03	26.85
1995	7002.5	2808.6	430.7	375.2	20.0	2062.40	666.95	7.90	61.97
199	8164.9	2898.8	429.4	417.3	21.1	2644.00	956.22	30.00	24.08
199	8982.4	3251.6	522.3	452.6	25.6	2819.60	1066.4	87.41	459.58
199	9463.0	3239.5	503.5	454.6	26.3	3126.76	1376.1	39.27	436.60
199	9913.6	3606.3	571.3	403.2	23.7	3534.98	1485.7	112.34	282.50
2000	10807.4	4742.9	660.0	407.2	22.4	3939.54	1558.28	186.24	561.93
2001	11757.3	5096.5	719.3	468.8	68.84	4254.06	1986.04	219.03	247.47
2002	12706.6	6207.7	754.6	527.4	27.5	5110.76	2324.29	138.47	41.07
2003	14182.7	8509.9	1012.3	535.1	29.0	6200.41	3114.18	114.27	299.62
200	16537.	11545.	1336.6	606.3	55.1	7865.53	4696.0	196.90	379.08

Source: "2004 China Statistical Yearbook" "Monetary Administration's Balance Sheet" compiled.

Table 2 China's economic growth rate, trade openness, foreign capital openness, financial openness, and openness to the outside world (unit: %)

Years	Economic growth rate	Trade openness	Openness of foreign investment	Financial openness	Openness to the outside world
1985	13.5	24.50	0.85	3.30	9.55
1986	8.9	22.43	0.79	3.38	8.87
1987	11.6	24.16	0.92	3.56	9.55
1988	11.3	23.80	1.01	3.66	9.49
1989	4.1	26.64	0.92	3.46	10.34
1990	3.8	32.30	1.11	5.62	13.01
1991	9.2	36.06	1.29	11.53	16.29
1992	14.2	38.06	3.08	7.96	16.36
1993	14.0	54.30	7.93	6.62	22.95
1994	13.1	49.49	6.59	13.30	23.13
1995	10.9	46.26	5.65	16.67	22.86
1996	10.0	40.80	5.37	18.41	21.53
1997	9.3	42.01	5.33	21.96	23.10
1998	7.8	39.55	5.08	24.52	23.05
1999	7.6	42.14	4.31	23.01	23.15
2000	8.4	50.00	3.99	23.24	25.74
2001	8.3	49.47	4.58	25.33	26.46
2002	9.1	54.97	4.37	23.45	27.60
2003	10.0	67.14	3.97	26.57	32.55
2004	10.1	77.89	4.00	31.59	37.83

Note: The data is calculated based on the above formula and the data in Table 1.

4. MODEL ESTABLISHMENT AND REGRESSION ANALYSIS

Based on the above-mentioned selection of external openness indicators, the study of the effects of economic liberalization on economic growth not only starts from foreign trade and foreign capital, but also adds a new perspective on financial capital, so the total import and export volume and actual foreign direct investment (The sum of FDI and GDP is a good measure of the degree of regional opening to the outside world, and the application of relevant statistical analysis software to empirically analyze the relationship. The expected conclusions will be a useful guide for a rational economic open policy.

The method of empirical analysis in this paper is mainly based on the economic growth model proposed by Solow in 1956. In this model, Solow assumes a two-factor production function:

$$\mathrm{Y} = \mathrm{F}(K, L) = AK^{\alpha}L^{\beta} \qquad (7)$$

Among them, K is capital, L is labor force, Y is output, and α and β are the output elasticities of capital and labor, respectively. From formula (7), we can see that in the Solow model, FDI and domestic capital are seen as homogeneous elements in the capital variable K, and Solow does not consider the impact of technological progress on output. In order to explain continued economic growth, it is necessary to consider external factors that have long increased factor productivity. Therefore, when the formula (7) incorporates the time factor, then:

$$\mathrm{Y} = \mathrm{F}(K, L, t) = e^{\gamma t}K^{\alpha}L^{\beta} \qquad (8)$$

In formula (8), where e is the base of the natural logarithm; t is the time; the rest is the same as the definition of formula (7). In fact, after the introduction of the time factor, all factors such as technological progress, changes in industrial structure, and institutional changes are attributed to the time coefficient. Therefore, it is called total factor productivity and is the growth rate of total factor productivity. Take the logarithm form of e in the formula (8), and add a random variable, you can get:

$$\mathrm{Y} = \mathrm{F}(K, L, t) = e^{\gamma t}K^{\alpha}L^{\beta} \qquad (9)$$

In fact, formula (9) assumes that domestic capital and foreign direct investment are homogeneous capital, which is inconsistent with China's actual economic conditions. Since the reform and opening up in 1978, the total capital used by China for investment has not only originated in China, but a considerable part of it has come from FDI. The inflow of FDI brings about advanced technology, management experience and institutional innovation in the investing countries. These intangible factors can be absorbed to a certain extent, thus affecting economic growth. Therefore, we cannot simply include FDI and domestic capital as homogeneous capital in the capital variable. Instead, FDI should be considered as a separate variable that affects China's economic growth.

Secondly, as the financial openness influencing the virtual economy, it is not considered in the formula (9). The impact of the accuracy of the results in the current period has entered the new century. The real economy in China has not only been greatly developed, but also the financial market is booming. The impact on economic growth has become increasingly apparent, so we will use it as a separate variable to influence economic growth. At the same time, the time factor t is introduced and changes accordingly.

For the sake of simplifying the analysis, assume that domestic capital is a homogeneous capital, that is, it can only be deployed in the form of capital in the domestic scope. FDI flows in the international scope in the form of capital and technology. It is different from domestic capital and is a heterogeneous capital. We can define the total capital level of a country as the weighted average of domestic capital and FDI. The mathematical form specifically states:

$$\mathrm{K} = K_d^{\lambda}K_f^{1-\lambda} \qquad (10)$$

$$\mathrm{Y} = \mathrm{f}\left(K^d, K^f, L, f, t\right) = e^{\gamma t_y + f t_f}\left(K^d\right)^{\alpha\lambda}\left(K^f\right)^{\alpha(1-\lambda)}L^{\beta} \qquad (11)$$

At the same time, in order to examine the role of foreign trade in China's economic growth, we include the import and export amount T as a variable in the production function, which is the output elasticity of T. Thus we obtain the final model:

$$\mathrm{Y} = \mathrm{f}\left(K^d, K^f, L, T, f, t\right) = e^{\gamma t_y + f t_f}\left(K^d\right)^{\alpha\lambda}\left(K^f\right)^{\alpha(1-\lambda)}L^{\beta}T^{\theta} \qquad (12)$$

Where K,K^d and K^f represent the total capital level, domestic capital, and FDI respectively. λ represents the weight of domestic capital in the total capital composition. After incorporating FDI as an input variable of the production function into the Cobb-Douglas production function, the financial openness f is also taken as a new factor, and it is tied to the time coefficient γ along with factors such as institutional changes, technological advances, and changes in the industrial structure, referred to as the new total factor productivity, and γ is the growth rate of total factor productivity, and f is the logarithmic form of the increase in financial openness, and is obtained by adding a random variable:

$$\mathrm{Ln}(Y_t) = \gamma t_y + f t_f + \alpha\lambda Ln\left(K_t^d\right) + \beta \mathrm{Ln}(L_t) + \alpha(1 - \lambda LnKtf + \theta LnTt + \mu t \quad (13)$$

Table 3 GDP, Domestic Capital Investment, Labor Force Input, FDI and Import and Export Volume from 1985 to 2004

Years	GDP (Billion)	K^d (Billion)	L(Ten thousand people)	K^f (Billion)	T(Ten thousand people)
1985	8964.4	2494.45	49873	48.75	2066.7
1986	10202.2	3056.03	51282	64.57	2580.4
1987	11962.5	3705.72	52783	85.98	30.84.2
1988	14928.3	4635.06	54334	118.74	3821.8
1989	16909.2	4282.64	55329	127.76	4155.9
1990	18547.9	4350.18	64749	166.82	5560.1
1991	21617.8	5346.11	65491	248.39	7225.8
1992	26638.1	7457.38	66152	622.72	9119.6
1993	34634.4	11472.17	66808	1600.13	11271.0
1994	46759.4	14117.25	67455	2925.69	20381.9
1995	58478.1	16885.88	68065	3133.38	23499.9
1996	67884.6	19444.45	68950	3469.10	24133.8
1997	74462.6	21189.39	69820	3751.72	26967.2
1998	78345.2	24642.24	70637	3763.93	26849.7
1999	82067.5	26516.98	71394	3337.73	29896.2
2000	89468.1	29547.18	72085	3370.55	39273.2
2001	97314.8	33333.4	73025	3880.09	42183.3
2002	105172.3	39134.37	73740	4365.54	51378.2
2003	117251.9	51138.00	74432	4428.61	70483.5
2004	159878.0	65054.65	75052	5018.35	95582.80

Sources of data: China Statistical Yearbook and the National Bureau of Statistics website

Using the statistical data of China's economy from 1985 to 2004 to regression (using Eviews software) for the model (13), the following estimation model can be obtained:

$$Ln(Y_t) = -2.1038t_\gamma + 0.9821t_f + 0.46107Ln(K^d) + 0.53327Ln(L) + 0.07538Ln(K^f) + 0.211Ln(T)$$

R^2 =0.9948 Adjusted- R^2 =0.9934 F-statistic=711.2 (P>F→<0.0001)

After estimating the parameters of the model, we found that after introducing the two variables of FDI and import and export volume, of the six factors affecting economic growth, at the 5% level of significance, only domestic capital investment and financial openness are significant. Each additional unit of domestic capital investment will increase GDP by 0.461 units. Every percentage point increase in financial openness will boost GDP growth by 0.98 units. The labor input is not significant. Each additional unit will increase the GDP by 0.533 units. In contrast, the amount of imports and exports and foreign direct investment are not significant. For each additional unit of the import and export volume, the GDP is increased by 0.211 units; for each additional unit of FDI, the GDP is increased by 0.075 units. Total factor productivity is very insignificant. However, from the whole model, the degree of quasi-optimization reached 0.99, and the model itself is also very significant. This shows that the six selected factors can explain China's economic growth situation well.

Through the above empirical analysis, the following basic conclusions can be drawn:

1. At present, domestic capital investment is still the primary factor in promoting China's economic growth. At the same time, China's financial level needs to be further improved in the first phase. In contrast, the current FDI amount and import and export volume are not the main factors affecting China's economic growth.

2. Foreign trade has a slightly stronger driving effect on economic growth than FDI and financial openness.

3. The inadequacy of China's trade, finance, and exchange rate systems has distort the prices of trade products to a certain extent, hinder the effective allocation of resources through domestic price trade through foreign trade and foreign trade, and have created inadequate domestic market development and utilization. The situation of excessive dependence on trade, in addition, the unsound system also increased the transaction costs in China's foreign trade and further reduced the efficiency of foreign trade. Therefore, we must further deepen structural reforms in trade, finance, and exchange rates, reduce foreign trade transaction costs, and improve the effective distribution mechanism of resources in the domestic and international markets. At the same time, we must pay full attention to the construction of domestic consumer markets, expand the proportion of domestic consumption in GDP, implement a strategy of winning with quality, and use precious resources for domestic economic construction.

5. POLICY IMPLICATIONS

This paper uses Eviews analysis software to examine the actual effects of economic opening on China's economic growth from the perspective of foreign investment and foreign trade and the level of China's financial liberalization. The following basic conclusions are reached: At present, domestic capital investment is still driving China's economic growth. The most important factor is that the contribution of labor input to China's economic growth is not significant, while TFP, FDI, and imports and exports have little effect on China's economic growth. Based on the empirical research in this paper, we can obtain the following policy implications by further analyzing the problems in China's open economy:

1. At present, domestic capital investment is still the primary factor that promotes China's economic growth, and total factor productivity, which embodies the factor of technological progress, even has a negative correlation with GDP. The role of investment in promoting economic growth based on inputs of production factors has the nature of diminishing marginal returns, and it requires long-term economic growth rates. From the perspective of economic growth strategy, we must attach importance to the role of technological progress (The role of technological progress factors in promoting economic growth has a marginal effect. The nature of income that is constant or increasing).

2. The current role of FDI in China's economic growth is very limited. As FDI promotes a country's economic growth mainly through capital formation, technology spillovers and structural adjustment effects, given that the quantitative effect of China's foreign investment has become quite significant, to further enhance FDI's catalytic role in China's economic growth, we can start with enhanced technology spillover effects, actively guide foreign investment, selectively attract multinational corporations with strong industry linkage effects and basic science and technology development orientation to make direct investment in China, and at the same time, enhance technological attraction and strengthen the regional advantages of high-tech industrial zones.

3. There is still much room for increase in the contribution rate of financial openness to China's economic growth. China must further increase its control and direction of financial policies. Due to the unsoundness of China's trade, finance, and

exchange rate systems, the prices of trade products have been distorted to a certain extent, and resources have been hindered from effectively integrating domestic trade and foreign trade through the price mechanism. This has led to a lack of development and utilization of the domestic market and an excessive dependence of economic growth on trade. In addition, the unsound system also increased the transaction costs in China's foreign trade and further reduced the efficiency of foreign trade. Therefore, we must further deepen structural reforms in trade, finance, and exchange rates, reduce foreign trade transaction costs, and improve the effective distribution mechanism of resources in the domestic and international markets. At the same time, we must pay full attention to the construction of domestic consumer markets, expand the proportion of domestic consumption in GDP, implement a strategy of winning with quality, and use precious resources for domestic economic construction.

In addition, the role of foreign trade in promoting China’s economic growth at this stage is not yet significant. Further analysis shows that: China's processing trade accounts for a large proportion, the quantitative expansion of trade structure and the adverse impact of foreign trade on China's independent technological innovation, makes foreign trade increase the efficiency of factor use through technology spillover effects and thus promote the economic growth of a country. The effect is also not ideal. Specifically, we can start from the following aspects to enhance the role of foreign trade in promoting economic growth in China. (1) Strengthen the industrial linkage effect of processing trade. We will vigorously develop the processing trade of high-tech products and promote the upgrading of processing trade. (2) Encourage comparative advantage enterprises to carry out the second venture while vigorously developing competitive advantage enterprises. (3) Take effective measures to protect and cultivate the independent innovation capabilities of Chinese enterprises and eliminate the negative impact of foreign trade on China's technological innovation

6. REFERENCE

[1]Harrison. A. Openness and Growth, A Time Series, Cross Country Analysis for Developing Countries [J]. Journal of Development Economics, 1996, (48): 419- 447.

[2]Bao Qun, et al. Trade Openness and Economic Growth: Theory and China's Empirical Study [J]. World Economy, 2003, (2): 10-18.

[3]He Zhengxia. Empirical Analysis of the Effect of Economic Opening on China's Economic Growth [J]. International Trade Issues, 2006,(10):17-21.

[4]Luo Zhongzhou. Empirical Analysis of Openness and Economic Growth in the Eastern Coastal Areas [J]. Finance and Economics Review,2007,(5):1-6.

[5]Lan Yisheng. An Empirical Analysis of Opening Degree and Regional Economic Growth in China [J]. Statistical Research, 2002,(2): 19-22.

[6]Chen Hong, Xu Yuqiang.Analysis of China's economic opening to the outside world [J].Heilongjiang Foreign Trade,2008,(7).

[7]Gao Hongye. Western Economics (Macro Part) [M]. Fourth Edition, March 2007. Renmin University of China Press, 2008, 684-724.

[8]Huang Xinfei, Tan Qiumei. Analysis of Trade Openness and Industrial Impact Mechanism of China's Economic Growth [J]. Academic Research.2007.10:82-84.

[9]Liang Xiaojuan. Empirical Analysis of the Relationship between Regional Economic Growth and Financial Development [J]. Journal of Henan Institute of Finance Management.2009.3:28-30.

[10]Zhang Cao. Empirical Analysis of the Relationship between Trade Liberalization and Economic Growth in China [J]. Modern Business.2007.3:205-206.

International Journal of Computer Applications Technology and Research
Volume 7–Issue 06, 208-214, 2018, ISSN:-2319–8656

The Effectiveness and Efficiency of Medical Images after Special Filtration from the View of Medical Specialist An Application on Khartoum Hospital-Sudan

Mohamed Y. Adam
Training and Community Service Center
King Saud University,
Riyadh, KSA

Dr. Mozamel M. Saeed
Collage of Science, Dept. of Computer Science
Prince Sattam Bin Abdul-Aziz University, KSA

Prof. Dr. Al Samani A. Ahmed
Computer Science
Al Neelain University
Khartoum, Sudan,

Abstract: There are many factors which have influences on the quality of medical images, so this paper gives a brief narration on the important techniques that produce acceptable quality to medical images. To ensure the validity of this techniques towards medical images, a questionnaire was designed and distributed to a number of doctors and professionals. The survey aims to assess the medical image specialists by regarding their point of views towards the impact of filtering medical images after processing using these techniques. MatLab package used to apply the techniques.

Keywords: Enhance, Logarithmic Transformations, Median Filtering, Histogram Equalization, Noise.

1. INTRODUCTION

Although modern digital cameras offer a great service to users in terms of facilitating the acquisition of images, but the user still needs to improve some of the images, which are unclear when taking the picture because of many reasons. The difference in light intensity from one location to another causes the instability of contrast, misty appearance and blurred colours. Therefore, in this paper different algorithms have been used to improve the images because they have a great effect on adjusting the lighting in dark images, clarifying their edges, clarifying their features and improving image quality[1]-[2].

In general, there are many things that make images to appear un normal such as: the resolution is poor , i.e. to increase the size of some details of the image, appearance of many types of noise and blurring caused by motion or lack of focus.

There are many techniques that are used to enhance an image including Logarithmic Transformations, Median Filtering, Histogram Equalization. The enhanced image obtained from the global area, histogram equalization will affect the intensity saturation in darkness area and whiteness areas. The colour image enhancement will be obtained by encoding the colour of red, green and blue to three different spectral images[1],[3]-[5].

2. IMAGE ENHANCEMENT TECHNIQUES

The image enhancement process aims to display certain features of the images for analysis or viewing. Medical images often suffer from one or more of the following defects:

- low resolution (in the spatial and spectral domains);
- high level of noise;
- low contrast;
- geometric deformations;
- presence of imaging artifacts.

These imperfections can be inherent to the imaging modality (e.g., X-rays offer low contrast for soft tissues, ultrasound produces very noisy images, and metallic implants will cause imaging artefacts in MRI) or the result of a deliberate trade-off during acquisition. For example, finer spatial sampling may be obtained through a longer acquisition time. This would also increase the probability of patient movement and thus blurring. In this paper, we will only be concerned with enhancing the medical images and we will not be interest in the challenging problem of designing optimal procedures for their acquisition.[7]-[9].

3. HISTOGRAM EQUALIZATION

The histogram equalization algorithm improves image contrast by converting image intensity values so that the resulting image graph is flat. Histogram Equalization is one of the most important techniques used to access high-quality images in colour scales in medical applications such as X-ray, MRI, CT scan. In order to identify diseases and correct diagnosis, these images require high resolution and colour variation.

To implement the histogram equalization on the image, you must obtain the probability density function and the cumulative density function of the image. This makes you able to calculate the number of pixels per colour in the image, and produce a cumulative total count. Then, by changing the output size, we can implement the histogram equation.

You can also implement the histogram equalization by reading the pixel density, looking for it in the pixel set and then measuring it accordingly to set the new pixel value. For the chart specification, a brief description is given here [6]-[7],[13].

Type the action that graphically displays a histogram. This will help you visualize the changes made by the graph balancing process and its specifications in an image. Below are MatLab commands to perform the above operations, as well as some other basic processes. In MatLab, we use the following commands.

```
img = imread (imageName);
```

```
imGray = rgb2gray(img);
myHist = imhist(imGray);
eqImage = imhisteq(imGray);
figure, imshow(imGray);
figure, imshow(eqImage);
```

In order to view the histogram of the image, you will supply two output arguments to the function for histogram equalization. The output arguments are the transformation function and the equalized image respectively. This is given as follows:

```
[eqImage, transfFunc] = imhisteq(imGray);
figure, plot(transfFunc);
figure, imshow(eqImage);
```

The second and third lines display the transformation function and the equalized image respectively.

In order to perform histogram specification, we use the same function as before (histeq), but include a specified histogram as the second argument as in the first line below

```
[imSpec, transSpec] = histeq(img, specPDF);
figure, imshow(imSpec), title('Matched Image');
[imSpecEq, transfSpecEq] = histeq(imSpec);
figure, plot(transfSpecEq), title('Specified Image
Histogram...');
```

The second line simply displays the matched image, the third line gets the histogram of the matched image, and the last line plots the histogram, which should match the specified PDF.

4. TYPES OF NOISE IN DIGITAL IMAGES

The image is usually composed of a set of discrete elements called pixels. The values of pixels that have an offset or motion create the noise in the image. Noise can appear through the cameras or through the image transfer process on different modes of transport.

If you want to restore the original image, noise must be removed. The noise in the image affects the application of the image processing algorithms so that the results are incorrect. There are need to filter the image before applying the algorithms.

5. MEDIAN FILTERING

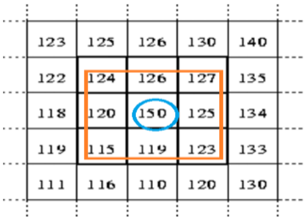
The most important filter from nonlinear filters is the median filter, which is used extensively in noise removal in a way that makes the picture retain its details. The median filter depends on where the noise is located and then is removed and replaced by the average of the neighborhood pixels, while the other points remain unchanged.

The median filter will be generated from the below equation where the pixel value of a point p is replaced by the result of median of pixels value of eight neighborhood of the point p.

$$g(p) = \text{median}\{f(p), \text{where } p \in N_8\ (p)\} \quad (1)$$

The median value will replace the central pixel according to brightness of the neighbouring pixels.

For example, we can calculate the median value of pixel neighbourhood of 150, as we can see in figure 1 below, and then is replaced with the median of surrounding pixels value that is 124. We use here A3x3 square neighbourhood



123	125	126	130	140
122	124	126	127	135
118	120	150	125	134
119	115	119	123	133
111	116	110	120	130

Figure (1): Neighbourhood values

As a result, the median value of the neighbourhood is: 124
The median filter advantages include:
– the contrast will not be changed because of the output values available from the present neighbourhood.
–Boundaries do not shift in median filtering that can happen with other filters.
–The outliers values will be removed because the median is less sensitive to the extreme values than the mean.

6. LOGARITHMIC TRANSFORMATIONS

If you want to increase the light intensity of the image, you have to use logarithmic transform. It is also often use to increase contrast for lower intensity values.

In MATLAB, the equation used to get the Logarithmic transform of image f is:
g = c*log(1 + double(f))

The constant c is usually used to scale the range of the log function to match the input domain. In this case c=255/log(1+255) for a uint8 image, or c=1/log(1+1) (~1.45) for a double image. It can also be used to further increase contrast—the higher the c, the brighter the image will appear. By using this way, the log function can produce values too bright to be displayed.

Notice that when c=5, the image is the brightest and you can see the radial lines on the inside of the medical image (these lines are barely viewable in the original because there is not enough contrast in the lower intensities).
The MATLAB code that created the image I when c=1, c=2, and c=3 is:

Notice the loss of detail in the bright regions where intensity values are clamped. Any values greater than one, produced from the scaling, are displayed as having a value of 1 (full intensity) and should be clamped. Clamping in MATLAB can be performed by the min(matrix, upper_bound),
Although logarithms may be calculated in different bases such as MATLAB's builtin log10, log2 and log (natural log), the resulting curve, when the range is scaled to match the domain, is the same for all bases. The shape of the curve is dependent instead on the range of values it is applied to. It is important to be aware of this effect if you plan to use logarithm transformations successfully.
Note that for domain [0, 1] the effects of the logarithm transform are barely noticeable, while for domain [0, 65535] the effect is extremely exaggerated. Also note that, unlike with linear scaling and clamping, gross detail is still visible in light areas.

7. The Results of the Questionnaire about the Application

To ensure the validity of these techniques, a questionnaire was designed and distributed to a number of doctors and professionals. In dealing with medical images, there were positive results to demonstrate the validity of these techniques for the purpose it was designed for.

Section 1: General information of interviewer

Table(1): sex

	Freq	%
male	13	65
female	7	35
Total	20	100

Table(2):age

	Freq	%
30 – less than 40	13	65
40- less than 50	7	35
Total	20	100

Table(3):job

	Freq	%
Doctor	7	35
Medical image technician	13	65
Total	20	100

Table(4):Years of Experience

	Freq	%
less than 5	5	25
5-less than 10	5	25
10-less than 20	10	50
Total	20	100

Table(5):Further education in the area of medical image processing techniques is important:

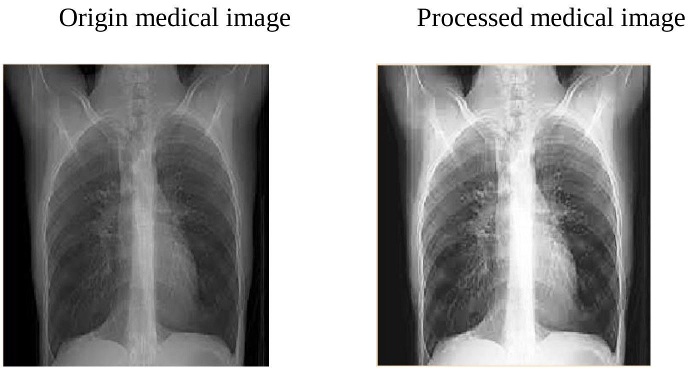
	Freq	%
Yes	17	85
No	3	15
Total	20	100

Table(6):In case the medical image is up normal: The treatment will be as follows:

	Freq	%
Repeating	6	30
Processing	14	70
Total	20	100

Section 2: The effects of medical image processing techniques using MatLab package:

a) Histogram equalization technique.



Origin medical image

Processed medical image

Figure (2): Applying histogram equalization

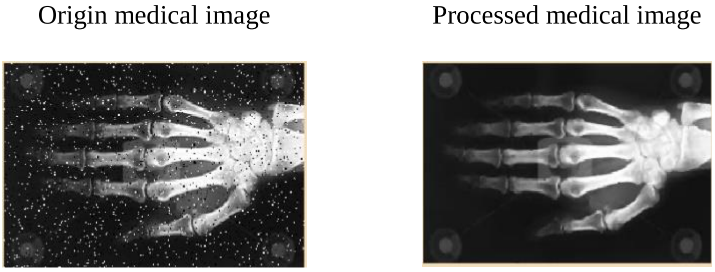
TABLE(7): THE RESULT OF HISTOGRAM EQUALIZATION TECHNIQUE QUESTIONS

Question		Answer			Average	Result
		Yes	Neutral	No		
1.The resolution in processed medical image is better than the origin medical image	Freq	19	1	0	1.05	Yes
	%	95.0	5.0	0		
2.The contrast Sensitivity in processed medical image is better than origin medical image.	Freq	13	7	0	1.35	Yes
	%	65	35	0		
3.Noises in processed medical image are less than in the origin medical image.	Freq	18	2	0	1.1	Yes
	%	90.0	10.0	0		
4.The blur in processed medical image is less than in the origin medical image.	Freq	13	7	0	1.35	Yes
	%	65	35	0		

5.In general the processed medical image (new medical image) is better than the origin medical image in order to diagnose	Freq	19	1	0	1.05	Yes
	%	95.0	5.0	0		
Total Result					1.12	Yes

Through table(7) above shows that most respondents agree that the resulting image after processing is better than the original image, depending on the trio scale likart

b) Median filter technique.


Origin medical image

Processed medical image

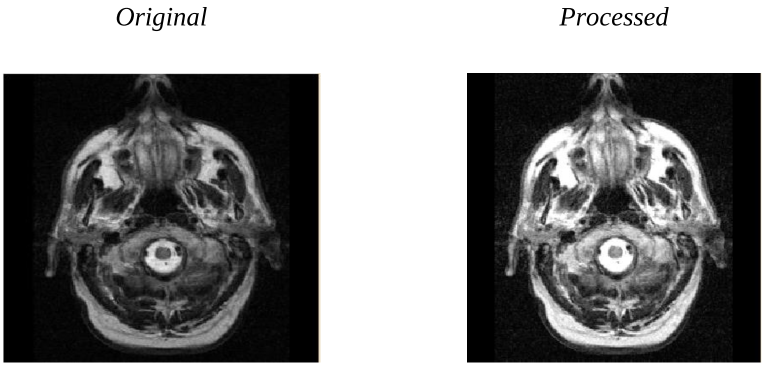
Figure (3): Applying Median filter technique

TABLE(8): THE RESULT OF MEDIAN FILTER TECHNIQUE QUESTIONS

Question		**Answer**			**Average**	**Result**
		Yes	**Neutral**	**No**		
1.The resolution in processed medical image is better than the origin medical image	Freq	20	0	0	1	Yes
	%	100	0	0		
2.The contrast Sensitivity in processed medical image is better than origin medical image.	Freq	13	7	0	1.35	Yes
	%	65	35	0		
3.Noises in processed medical image are less than in the origin medical image.	Freq	18	2	0	1.1	Yes
	%	90.0	10.0	0		
4.The blur in processed medical image is less than in the origin medical image.	Freq	13	7	0	1.35	Yes
	%	65	35	0		
5.In general the processed medical image (new medical image) is better than the origin medical image in order to diagnose	Freq	20	0	0	1	Yes
	%	100	0	0		
Total Result					1.12	Yes

Through table(8) above shows that most respondents agree that the resulting image after processing is better than the original image, depending on the trio scale likart.

c) Logarithm transformation


Original

Processed

Figure(4): Applying Logarithm transformation

TABLE(9): THE RESULT OF LOGARITHM TRANSFORMATION TECHNIQUE QUESTIONS

Question		Answer			Average	Result
		Yes	Neutral	No		
1.The resolution in processed medical image is better than the origin medical image	Freq	18	2	0	1.1	Yes
	%	90.0	10.0	0		
2.The contrast Sensitivity in processed medical image is better than origin medical image.	Freq	13	7	0	1.35	Yes
	%	65	35	0		
3.Noises in processed medical image are less than in the origin medical image.	Freq	18	2	0	1.1	Yes
	%	90.0	10.0	0		
4.The blur in processed medical image is less than in the origin medical image.	Freq	13	7	0	1.35	Yes
	%	65	35	0		
5.In general the processed medical image (new medical image) is better than the origin medical image in order to diagnose	Freq	20	0	0	1	Yes
	%	100	0	0		
Total Result					1.12	Yes

Through table(9) above shows that most respondents agree that the resulting image after processing is better than the original image, depending on the trio scale likart.

8. CONCLUSION

The main goal of this study is to build an application to enhance medical images using effective enhancement algorithms including histogram equalization, Median filter technique, Logarithm transformation. The developed application has been tested and deployed by the staff of Khartoum hospital. According to the view of medical specialists, these algorithms are powerful method for image enhancement and they will increase the contrast of image. The results are plotted in the above tables.

The proposed methods have been implemented by the strongest and most popular program in Computational Libraries (MATLAB).

9. REFERENCES

[1] Achmad, B., M.M. Mustafa and A. Hussain,"Warped optical-flow inter-frame reconstruction for ultrasound image enhancement.".jcssp.1532.1540DOI:10.3844, .2011.

[2] Charles Henry Brase & Corrinne Pellillo Brase, Understandable Statistics: Concepts and Methods, 2014

[3] D.A. Forsyth and J. Ponce, Computer Vision – A Modern Approach, Prentice Hall, 2003

[4] Gregory d Abram, Parallel image generation, with anti-aliasing and texturing. University of North Crolina Chapel Hill NC27599-3175

[5] Image Processing using Matlab. Second Edition. United States of America. Gatesmark Publishing.

[6] Mohamed Y. Adam, Mozamel M. Saeed and Al Samani A. Ahmed, "Medical Image Enhancement Application Using Histogram Equalization in Computational Libraries", International Journal of Computer Science and Telecommunications, Volume 6, Issue 1, January 2015, pp 7-12, http://www.ijcst.org/Volume6/Issue1/p2_6_1.pdf

[7] Mohamed Y. Adam, Mozamel M. Saeed and Al Samani A. Ahmed, " THE EFFECT OF IMPLEMENTING OF NONLINEAR FILTERS FOR ENHANCING MEDICAL IMAGES USING MATLAB", International Journal of Computer Science & Information Technology (IJCSIT) Vol 7, No 6, December 2015
http:// www .aircconline.com/ijcsit/V7N6/7615ijcsit05.pdf

[8] Mrozek, Bogumiła – Mrozek, Zbigniew., Matlab 6; Poradnikużytkownika, 2010.

[9] Nayan Patel, Abhishek Shah, MayurMistry, "Astudy of Digital Image Filtering Techniques in Spatial Image Processing", International Conference on Convergence of Technology – 2014

[10] Ozimek and Agnieszka b., Digital image processing. Materials from Lecture no. 7. Cracow. Cracow University of Technology, 2010

[11] Processing using Matlab. Second Edition. United States of America. Gatesmark Publishing.

[12] Rafael C. Gonzalez and Richard E. Woods, Digital image processing, Third edition, Prentice Hall, 2008

[13] Rakesh M.R, Ajeya B, Mohan A.R,"Hybrid Median Filter for Impulse Noise Removal of an Image in Image Restoration",International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 10, October 2013, http://ijareeie.com/upload/2013/october/8QHybrid.pdf

[14] Sigurd angenent, eric pichon, and allen tannenbaum, "mathematical methods in medical image processing", http://www.math.wisc.edu/~angenent/preprints/medicalBAMS.pdf

[15] The MarthWorks, Image Processing Toolbox 6 User's Guide. United States of America. The MathWorks, 2009.

[16] Warner and David , Digital Image processing – an analytic approach Prentice Hall ,2003

10. Appendix:

1.Code:

```
a.function pushbutton4_Callback(hObject, eventdata,
handles)
global im
GIm=im;
numofpixels=size(GIm,1)*size(GIm,2);
HIm=uint8(zeros(size(GIm,1),size(GIm,2)));
freq=zeros(256,1);
probf=zeros(256,1);
probc=zeros(256,1);
cum=zeros(256,1);
output=zeros(256,1);
for i=1:size(GIm,1)
    for j=1:size(GIm,2)
        value=GIm(i,j);
        freq(value+1)=freq(value+1)+1;
probf(value+1)=freq(value+1)/numofpixels;
```

```
   end
end
sum=0;
no_bins=255;
for i=1:size(probf)
  sum=sum+freq(i);
  cum(i)=sum;
  probc(i)=cum(i)/numofpixels;
  output(i)=round(probc(i)*no_bins);
end
for i=1:size(GIm,1)
   for j=1:size(GIm,2)
         HIm(i,j)=output(GIm(i,j)+1);
   end
end
GIm=rgb2gray(GIm);
axes(handles.axes4);imshow(HIm);title('Image After
Histogram equalization');
 b.k = imread('ches.jpg');
k = rgb2gray(k);
subplot(2,2,1);imshow(k),title('ORIGINAL IMAGE');
g = imnoise(k,'salt& pepper');
subplot(2,2,2);
imshow(g);title('Noisy Image')
B=zeros(size(g));
modifyA=padarray(g,[1 1]);
      a=[1:3]';
      b=[1:3]';
for x= 1:size(modifyA,1)-2
for y=1:size(modifyA,2)-2
   mm=reshape(modifyA(x+a-1,y+b-1),[],1);
      B(x,y)=min(mm);
end
end
B=uint8(B);
subplot(2,2,3),imshow(B),title('IMAGE AFTER MIN
FILTERING');
k = imread('ches.jpg');
k = rgb2gray(k);
subplot(2,2,1);imshow(k),title('ORIGINAL IMAGE');
n = imnoise(k,'salt& pepper');
subplot(2,2,2);
imshow(n);title('Noisy Image')
B=zeros(size(n));
modifyA=padarray(n,[1 1]);
      a=[1:3]';
      b=[1:3]';
for x= 1:size(modifyA,1)-2
for y=1:size(modifyA,2)-2
     mx=reshape(modifyA(x+a-1,y+b-1),[],1);
     B(x,y)=max(mx);
end
end
B=uint8(B);
subplot(2,2,3);imshow(B),title('IMAGE AFTER MAX
FILTERING');
I = imread('ches.jpg');
subplot(2,2,1);
imshow(I);
title('Original Image')
J = imnoise(I,'salt& pepper',0.02);
subplot(2,2,2)
subimage(J)
title('Noisy Image')
L = medfilt2(J,[3 3]);
subplot(2,2,3);
imshow(L)
title('Noisy Image filtered by median filter')
c.function logarithm_Callback(hObject, eventdata,
handles)
global im
I2=im2double(im);
%J=1*log(1+I2);
J2=2*log(1+I2);
%J3=5*log(1+I2);
axes(handles.axes2);
imshow(J2)
```

2. The questionnaire distributed

Section 1: General information

1. Name (optional): …………………………………………

2. Sex: Male [ ] Female [ ]

3. Age: 20- less than 30 [ ] 30 – less than 40 [ ] 40- less than 50 [ ] 50 or more [ ]

4. Job: Medical Doctor [ ] Medical image Technician [ ]

5. Years of Experience: less than 5[ ] 5-less than10 [ ] 10-less than20 [ ] 20 or more [ ]

6. Further education in the area of medical image processing techniques is important:

Yes [ ] No [ ]

7. In case the medical image is up normal: The treatment will be as follows:

-Repeating the medical image [ ]

-Processing the medical image [ ]

Section 2: The effects of medical image processing techniques using MatLab package:

1. **Histogram equalization technique.**

Please put the sign correct (√) in the suitable place on the table Below:

	Yes	Neutral	No
1. The resolution in processed medical image is better than the origin medical image			
2. The contrast Sensitivity in processed medical image is better than origin medical image.			
3. Noises in processed medical image are less than in the origin medical image.			
4. The blur in processed medical image is less than in the origin medical image.			
5. In general the processed medical image (new medical image) is better than the origin medical image in order to diagnose			

2. **Median filter technique.**

Please put the sign correct (√) in the suitable place on the table Below:

	Yes	Neutral	No
1. The resolution in processed medical image is better than the origin medical image			
2. The contrast Sensitivity in processed medical image is better than origin medical image.			
3. Noises in processed medical image are less than in the origin medical image.			
4. The blur in processed medical image is less than in the origin medical image.			
5. In general the processed medical image (new medical image) is better than the origin medical image in order to diagnose			

3. **Logarithm transformation**

Please put the sign correct (√) in the suitable place on the table Below:

	Yes	Neutral	No
The resolution in processed medical image is better than the origin medical image			
The contrast sensitivity in processed medical image is better than origin medical image.			
Noises in processed medical image are less than in the origin medical image.			
The blur in processed medical image is less than in the origin medical image.			
In general the processed medical image (new medical image) is better than the origin medical image.			

International Journal of Computer Applications Technology and Research
Volume 7–Issue 6, 215-220, 2018, ISSN:-2319-8656

Employment Recommendation System using Matching, Collaborative Filtering and Content Based Recommendation

Roshan G. Belsare
Department of Computer
Science and Engineering
PRMIT&R, Badnera
Maharashtra, India

Dr. V. M. Deshmukh
Department of Computer
Science and Engineering
PRMIT&R, Badnera
Maharashtra, India

Abstract: The tremendous growth of both information and usage has led to a so-called information overload problem in which users are finding it increasingly difficult to locate the right information at the right time Thus huge amount of information and easy access to it make recommender systems unavoidable [1]. We use recommender system every day without realizing it and without knowing what exactly happens. Recommender systems have changed the way people find products, information, and even other people. They study patterns of behavior to know what someone will prefer from among a collection of things he/she has never experienced. Benefits of recommender systems to the businesses using them include: The ability to offer unique personalized service for the customer, Increase trust and customer loyalty, Increase sales, click-through rates, conversions, etc., Opportunities for promotion, persuasion and Obtain more knowledge about customers. Recommender systems are software tools and techniques providing suggestions for items to be of use to a user. Job recommender systems are desired to attain a high level of accuracy while making the predictions which are relevant to the customer, as it becomes a very tedious task to explore thousands of jobs, posted on the web, periodically. Although a lot of job recommender systems[2] exist that use different strategies, here efforts have been put to make the job recommendations on the basis of candidates profile matching as well as preserving candidates job behavior or preferences. Firstly, the rules predicting the general preferences of the different user groups are mined. Then the job recommendations to the target candidate are made on the basis of content based matching as well as candidate preferences, which are preserved either in the form of mined rules or obtained by candidates own applied job history.

Keywords: Recommendation System, Collaborative Filtering, Content Based Recommendation, Cosine based Similarity, Hybrid Recommendation, Information Retrieval.

1. INTRODUCTION

In recent years, the volume of data present online has grown exponentially. A major portion of this data is related to internet-based different platforms. The evaluation of such data and/or the extraction of information is difficult due to its huge volume. It is cumbersome for an individual or an organization to obtain the desired results in a timely manner. Hiring the right talent is a challenge faced by all companies. This challenge is amplified by the high volume of applicants if the business is labor intensive, growing and faces high attrition rates. One example of such a business is IT services run out of growth markets. In a typical services organization, professionals with varied technical skills and business domain expertise are hired and assigned to projects to solve customer problems. In the past few years, IT services including consulting, software development, technical support and IT outsourcing has witnessed explosive growth, especially in growth markets like India and China. For in-stance, according to a NASSCOM (National Association of Software and Services Companies of India) study, the total number of IT and IT enabled services professionals in India has grown from 284000 in 1999-2000 to over 1 million in 2004-2005 [12]. More recent estimates suggest that this industry employs more than 2 million professionals in India alone. For organizations in the IT Services business, growth in business is synonymous with growth in the number of employees and recruitment is a key function. Hiring large number of IT professionals in growth markets poses unique challenges. Most countries in growth markets have large populations of qualified technical people who all aspire to be part of the explosive growth in the IT Services industries. Thus, a job posting for a Java programmer can easily attract many tens of thousands of applications in a few weeks. Most IT Services companies are inundated with hundreds of thousands of applicants. For example, Infosys, one of the largest IT Outsourcing companies in India, received more than 1.3 million job applications in 2009. However, only 1% of them were hired. To give the context for work, consider a typical recruitment process. This is illustrated in Figure. The process starts when a business unit decides to hire employees to meet its business objectives. The business unit creates a job profile that specifies the role, job category, essential skills, location of the opening and a brief job description detailing the nature of work. It might also specify the total work experience that the prospective employee should possess, along with the desired experience level for each skill. The job openings are advertised through multiple channels like on {line job portals, newspaper advertisements, etc. Candidates who are interested to apply for the job opening upload their profile through a designated web-site. The website typically provides an on{line form where the candidate enters details about her application like personal information,[5] education and experience details, skills, etc. We call this Candidate Meta {data. The candidates can also upload their resumes through the website. The objective of allowing the candidate to enter meta {data in an on{ line form is to capture the information in a more structured format to facilitate automated analysis. However, real life experience suggests that most candidates do not specify a lot of information in the on [6]{line forms and hence Candidate Meta{data is often incomplete} Once the applications of prospective candidates are received, they are subjected to careful scrutiny by a set of dedicated screeners.

This screening process[17] is crucial because it directly affects the quality of the intake and hence, the company profits.

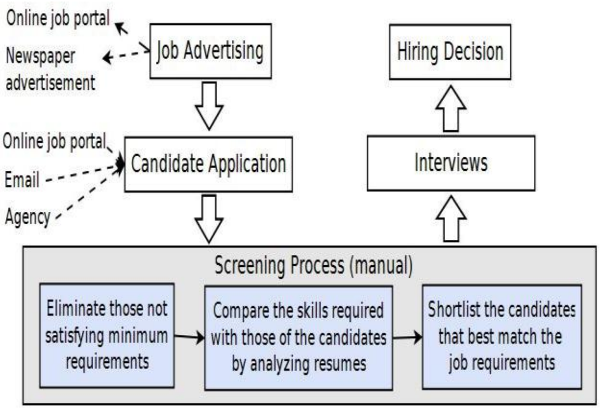

Online job portal
Newspaper advertisement
Job Advertising
Hiring Decision
Online job portal
Email
Agency
Candidate Application
Interviews
Screening Process (manual)
Eliminate those not satisfying minimum requirements
Compare the skills required with those of the candidates by analyzing resumes
Shortlist the candidates that best match the job requirements

Figure 1.1: Recruitment process with manual screening

The screeners typically proceed as below [18]:

1. Understand the requirement for the job opening, in terms of the skills that are mandatory and those that are optional but preferable, the experience criteria if any, preference for the location of the candidate etc. Also, note the kind of work that will be performed as part of the job role.

2. Look through each of the applications, and reject those who do not have the minimum years of experience or the skills required for the job.

3. Out of the remaining candidates, and the best match for the job. This requires the recruiter to read the resume in detail and compare it with the job pro le. Since the number of candidates who can be interviewed is limited, the recruiter has to make a relative judgment on the candidates.

The top few candidates, who are shortlisted during the screening, undergo further evaluation in the form of interviews, written tests, group discussions etc. The feedback from these evaluation processes is used to make the final hiring decision.

Recommender Systems (RS) present an automated and efficient solution to this problem. Recommender systems analyze the user profile/behavior [4][22] and suggest products/services relative to the user's interests. The recommender system technology plays an important role in various e-commerce applications by helping individuals to find right items in a large option space, which match their interests.

The problem of recommending jobs to users is fundamentally different from traditional recommendation system problems such as recommending books, products, or movies to users. While the entire above have a common objective to maximize the engagement rate of the users, one key difference is that a job posting is typically meant to hire one or a few employees only, whereas the same book, product, or movie could be potentially recommended to hundreds of thousands of users for consumption. [13]

Ideal job recommendation system would need to recommend the most relevant jobs to users.

A job recommender system is expected to provide recommendations in 2 ways: firstly recommending most eligible candidates for the specified job, to the recruiters and secondly, recommending jobs to the aspiring candidates according to their matching profiles. The focus of this paper is the second part only i.e. to recommend jobs to the candidates according to their matching profiles.

2. Objectives

The aim of recommender systems is to assist users in finding their way through huge databases and catalogues, by filtering and suggesting relevant items taking into account or inferring the users‟ preferences (i.e., tastes, interests, or priorities). Based on this objectives for Job Recommendation systems are

1. Study of matching technique to match the suitable candidate for job position.
2. Use of Collaborative Filtering to find best suitable match.
3. Use content-based approach that takes into consideration an organization's needs and the skills of candidate.
4. Comparison of matching technique with Collaborative Filtering and Content Based Recommendation.

3. Literature Review

Job Recommendation work resides in the domain of online recommender systems, which are widely adopted across many web applications, e.g., movie recommendations [12], e-commerce item recommendations [13], job recommendations [14] and so forth, where authors mainly concentrate on the relevance retrieval and ranking aspects of the recommendation system. There is insightful research and modeling of the hiring processes within job marketplaces. Such research includes work related to estimation of employee reputation for optimal hiring decisions [15], as well as work related to ranking and relevance aspects of job matching in labor marketplaces [16]. There has been work related to the theory of optimal hiring process, e.g., on the problem of finding the right hire for a job (the hiring problem), as well as on the classical secretary problem, where a growing company continuously interviews and decides whether to hire applicants [17,18]. Authors of [19] investigated job marketplace as a two-sided matching market using locally stable matching algorithms for solving the problem of finding a new job using social contacts. RS can be treated as one of the most efficient tools for business, aimed directly at increasing revenue and profitability as well as optimizing current product portfolio.

4. Proposed System

4.1 Brief Description

Proposed job recommendation system in this paper would take the input from tab separated files (TSV) to analyze the dataset. After analysis of the data the data would be transformed in the form of matrix by applying item-item collaborative filtering or user-user collaborative filtering. Once we have the data organized in the form of matrix we can apply different algorithms and machine learning techniques to find out the most suitable job for each user based on his or her analysis.

The central task of this challenge is to predict those job postings (items) that the user will interact with. Given a user, a job recommendation system should predict that job position that is likely to be relevant to the user. Proposed system makes use of Collaborative Filtering using cosine similarity and Content based recommendations using FirstText library.

4.2 Proposed System Design

The proposed Employment Recommendation System is divided mainly into four modules: Data acquisition module, Transformation Module, Computation module and Recommendation Module

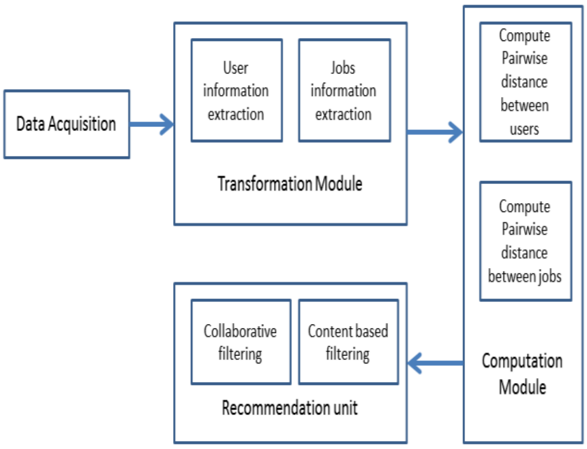

Data Acquisition
User information extraction
Jobs information extraction
Transformation Module
Compute Pairwise distance between users
Compute Pairwise distance between jobs
Computation Module
Collaborative filtering
Content based filtering
Recommendation unit

Fig.4.2.1 Employment Recommendation System Modules

Data acquisition module: In data acquisition module, data from different users are collected and stored in dataset. These data consist of use profile, skills, past activities, browsing history. For this purpose we are storing data in tab separated values (tsv) format files.

Transformation Module: Transformation stage deals with the different processes such as general information extraction and detail information extraction as per the requirement. If collaborative approach needs to be used in that case we are going to transform data present in dataset into the user-user and user-item matrices to be analyzed for recommendation.

Computation module: Computation module deals with the calculation part of the system It mainly consists of two parts - Data filtering and Result set generation. In the data filtering we need to analyze the similarity between item-item and user-item.

Recommendation Module: The last stage is the Recommendation unit where recommendations to the users are made depending upon the filtered results set from computation module.

Matching based Recommendation design:

In the matching between people and jobs, the content is the personal information [20] and their job desires for people while for jobs, it's the job description posted by recruiters, even including the background description of enterprises. In our case we have considered two parameters namely City and State for matching and recommending jobs to the users based on these two parameters based on the assumption that the user will prefer first the popular jobs in his city and followed by his state. Steps for recommending jobs to the users are as below:

Step 1: Read input data about all the jobs available from jobs.tsv file

Step2: Read the data about job applications from apps.tsv file

Step 3: Sort the jobs based on popularity

Step 4: Create a directory of available jobs based on city, state and popularity

Step 5: Predict the top 10 matching jobs for each user

Content Based Recommendation design:

The principle of a content-based [3] recommender [9][10] is to suggest items that have similar content to ones the target user prefers. The process of content-based recommender is selecting the same feature type and comparing them by calculating their similarity for people and jobs [23]. The recommendatory result is a list of job positions or candidates sorted by the similarity index. In short, the two key components of content-based recommender [21] are feature selection and similarity calculation. During selecting feature, not only it's need to select the common feature but also considering its influence on recommendation according to the target user's preferences or the scientific analysis in the job recruiting market. Then the selected features should be represented in an appropriate form, for instance vector space model and their similarity can be calculated.

Our content-based recommender system uses the Vector Space Model (VSM) [32]. In the VSM, each job is represented by a vector in an n-dimensional space, where each dimension corresponds to a textual feature from the overall feature-set of the job collection.

Let J = (J1,J2,...,JN) denote a set of jobs and F = (f1,f2,...,fm) be the feature set. Formally, every job Ja is represented as a vector of feature weights, where each weight indicates the degree of association between the job and the feature:

$J_a = (w_1^a, w_2^a, \ldots, w_m^a)$, where w_i^a is the weight of feature f_i for job J_a.

For feature weighting, we have used FirstText Deep learning library from Facebook.

Job Similarity Assessment: For a given feature representation [33], the similarity between two jobs, Ja and Jb is computed by the cosine similarity between their vector space representations, as follows:

$$sim(J_a, J_b) = \frac{\sum_{i=1}^{m} w_i^a \times w_i^b}{\sqrt{\sum_{i=1}^{m}(w_i^a)^2} \times \sqrt{\sum_{i=1}^{m}(w_i^b)^2}} \quad ,$$

where w_i^a and w_i^b are the weights of the feature f_i in job J_a and job J_b respectively.

Collaborative Filtering based recommendation Design:

The Collaborative filtering is one of the most successful approaches for building recommender systems. In Collaborative filtering important step of achieving matching jobs and people is calculating the similarity or relevance based on their profiles. Several common similarity calculation measures, namely, Constrained Pearson Correlation, Pearson Correlation, Spearman Rank Correlation, Cosine and Mean Squared Differences in the recommender system. For example, with explicit rating information we can measure the similarity between two users or two jobs. We are going to use Cosine similarity and Mean Square differences in this employment recommendation system. The similarity measure is the measure of how much alike two data objects are. Similarity measure in a data mining context is a distance with dimensions representing features of the objects. If this distance is small, it will be the high degree of similarity where large distance will be the low degree of similarity. The similarity is subjective and is highly dependent on the domain and application. In this technique analysis of the user-item matrix to discover relations between different users or items and use them to compute the recommendations [35]. In our CF algorithm after the k most similar applicants have been identified, their corresponding rows in the jobs/applicants ratings matrix R are aggregated to identify the set of jobs, J, rated by the group together with their ratings. We predict ratings for the current applicant. Then, we recommend the top-N jobs depending on their ratings.[36]

Computation process:

Cosine based Similarity:

Moving forward to this similarity measure, any two things are taken as two items in the s dimensional client-space. The concept of angle is used here to calculate the similarity among the different items. The similarity between the two items [7] is calculated by finding out the cosine of the angle between the taken any two items. Formally, in the n × m ratings matrix (that is user-item matrix) , similarity between any items let suppose that we are taking the arbitrarily items i and j, denoted by

$$sim(i,j) = \cos(i,j) = \frac{i.j}{\|i\|^{\wedge 2} * \|j\|^{\wedge 2}}$$

Steps for generation of recommendation list using different similarity measures are as follows:

Phase 1

Input: User-item matrix n*m that is R and k that indicates the count of job to job similarities that will be stored for each job.

Result: m*m matrix M

Step 1: The user-job matrix is taken and job to job similarity is calculated using the different similarity measure.

Step 2: The value of M (i, j) is compared with the k most similar jobs. If it’s same then it is left as the value is else it is made zero.

This is what we get the output matrix M, which will be used in the next phase of the algorithm.

Phase 2

Input: The output matrix M from the previous phase, the matrix m*1 U which store the products that has been purchased beforehand by the users, and the variable N that specifies the number of items that will be recommended to the users.

Output: m*1 matrix x that stores number of jobs to be recommended. Its non-zero value indicates that the jobs that is in top n and is recommended to its users.

5. Comparison between recommendation approaches

All three approaches proposed in this paper that is matching based recommendation, collaborative based recommendation and content based recommendation have recommended / predicted different set of top 10 jobs recommendation for each user. If we compare all three approaches it is clear that matching algorithm will produce fastest results/recommendation but the recommendations might not be that much useful to the jobseekers as it only tries to match jobs based on certain parameters and does not contain any personalized information[23]. On the other hand both collaborative and content based recommendations will predict/recommend jobs that are personalized for that user and hence might be much relevant to the users. It is also clear that Collaborative filtering arrives at a recommendation that's based on a model of prior user behavior.[34] The model can be constructed solely from a single user's behavior or more effectively also from the behavior of other users who have similar traits. When it takes other users' behavior into account, collaborative filtering uses group knowledge to form a recommendation based on like users. Comparison of these three approaches is summarized in below table:

	Matching based recommendation	Collaborative recommendation	Content based recommendation
Parameters Considered	City, State	Similarity between jobs	Job description is used for content based analysis
Personalized recommendations	No	Yes	Yes
Execution speed for generating recommendation	Fast	Slow	Very slow
Relevance to the user	May be	Based on ranking of similar jobs	Based on ranking of past behavior
Issues faced	No personalization	Cold start problem	Over specialization

Table 5.1: Comparison of different recommendation approaches

6. Results

The results of Cosine similarity based Collaborative filtering are promising in comparison to the results of matching technique. We have generated recommendations of collaborative filtering for different input data set and results generated are found to be more appropriate with the increased data set size.

Comparison of RMSE (Root Mean Square Error) for different size of input dataset for Collaborative filtering is shown in figure 6.1

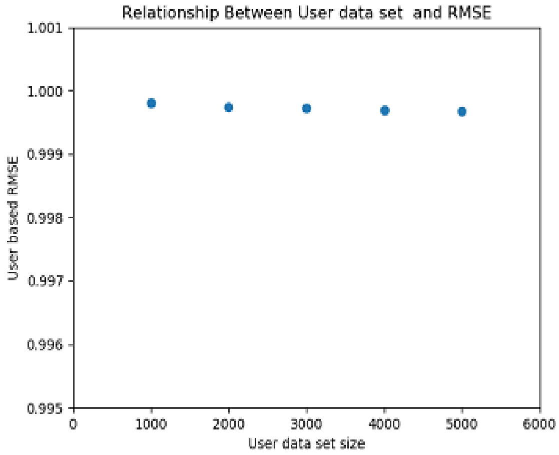


Figure. 6.1 Relationship between user data set and RMSE

Form the figure 6.1 it is clear that as we increase the size of dataset there is decrease in Root Mean square error and hence accuracy of prediction made by recommendation system is increasing with increase in input data samples

Comparison of User based RMSE and Job based RMSE is compared and results are shown in Figure 6.2

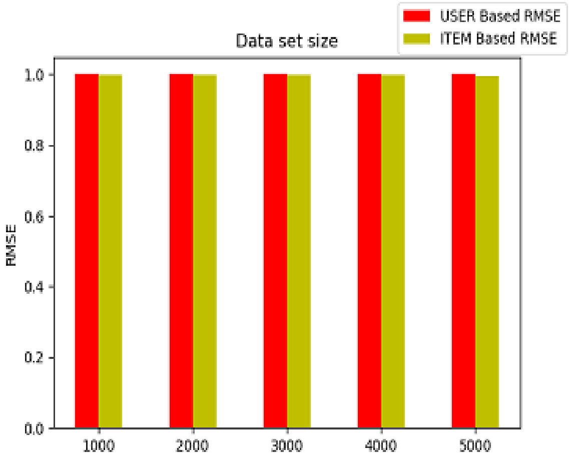


Figure 6.2 Bar chart of User based RMSE, item based RMSE with Data set size

Both user based RMSE and item based RMSE performance is shown in graph for increasing size of input data set and RMSE is decreasing with increase in input data set and accuracy of prediction by recommendation system is increasing with increase in input data size.

7. Conclusion

We have designed the employment recommendation system using different recommendation techniques like simple matching and then collaborative and content based filtering. Similarity metrics are used to calculate how much similar all the items are to each other in the matrix. We implemented the algorithm and get the result accordingly. Comparison is made between different recommendation approaches. While the Matching techniques is the simplest one but recommendations that are generated might not be that useful to the jobseekers as they are not personalized. Both collaborative filtering and content based recommendation have generated the personalized recommendations and hence they are more useful to the jobseekers. But collaborative filtering suffers from cold start problem while content based recommendation might generate too specific results.

8. Future Scope

In this field of job recommendation system there is a large scope for future work such as:

- By using different similarity measure we can see which gives the most accurate answer when compared with the other similarity measures.
- If we take into consideration the recommendation of the recommender system in contrast with the real life preferences we can compare their mean absolute error.
- We can consider large number of parameters by giving them associated weights for more accurate Content Based recommendation
- Different Content based approaches can be compared like TF-IDF, Word to Vec with FirstText[8]
- Hybrid recommdations [11] can be generated either by combining the approaches or by combining the outputs generated by Content based and Collaborative recommdations.
- We can perform natural language processing to extract information from jobseekers resume and then recommending him the jobs

9. REFERENCES

[1] K. Wei, J. Huang, and S. Fu. A survey of e-commerce recommender systems. In 2007 International Conference on Service Systems and Service Management, pages 1-5, June 2007.

[2] Chenrui Zhang , Xueqi Cheng An Ensemble Method for Job Recommender Systems. RecSys Challenge '16, September 15 2016, Boston, MA, USA 2016 ACM

[3] N. D. Almalis, G. A. Tsihrintzis and N. Karagiannis, "A content based approach for recommending personnel for job positions," IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications, Chania, 2014, pp. 45-49.

[4] M. Balabanovic, and Y. Shoham, "Fab: Content-based, Collaborative Recommendation. Communications of the ACM," vol. 40, no. 3, pp. 66- 72, 1997.

[5] M. Ramezani, L. Bergman, R. Thompson, R Burke, and B. Mobasher, "Selecting and Applying Recommendation

Technology," In proceedings of International Workshop on Recommendation and Collaboration in Conjuction with International ACM on Intelligence User Interface, 2008.

[6] BadulSarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms," Proceedings of the 10th International Conference of World Wide Web, pp. 285-295, 2001.

[7] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Computing, vol. 7, no. 1, pp. 76–80, 2003

[8] D. Mladenic, "Text-learning and Related Intelligent Agents: A Survey," IEEE Intelligent Systems, vol. 14, no. 4, pp. 44–54, 1999.

[9] RJ. Mooney and L. Roy, "Content-Based Book Recommending Using Learning for Text Categorization," in Proceedings of DL "00: Proceedings of the Fifth ACM Conference on Digital Libraries, New York, NY, ACM pp. 195-204, 2000.

[10] N. D. Almalis, G. A. Tsihrintzis and N. Karagiannis, "A content based approach for recommending personnel for job positions," IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications, Chania, 2014, pp. 45-49.

[11] Toon De Pessemier, Kris Vanhecke, and Luc Martens. 2016. A scalable, high-performance Algorithm for hybrid job recommendations. In Proceedings of the Recommender Systems Challenge (RecSys Challenge '16). ACM, New York, NY, USA, Article 5, 4 pages. DOI: https://doi.org/10.1145/2987538.2987539

[12] Greg Linden, Brent Smith, and Jeremy York. 2003. Amazon.Com Recommendations:Item-to-Item Collaborative Filtering. IEEE Internet Computing 7, 1 (2003), 76–80. https://doi.org/10.1109/MIC.2003.1167344

[13] Viet Ha-Thuc, Ye Xu, Satya Pradeep Kanduri, Xianren Wu, Vijay Dialani, Yan Yan, Abhishek Gupta, and Shakti Sinha. 2016. Search by Ideal Candidates: Next Generation of Talent Search at LinkedIn. In WWW. https://doi.org/10.1145/2872518.2890549

[14] Marios Kokkodis, Panagiotis Papadimitriou, and Panagiotis G. Ipeirotis. 2015. Hiring Behavior Models for Online Labor Markets. In WSDM. https://doi.org/10.1145/2684822.2685299

[15] Jia Li, Dhruv Arya, Viet Ha-Thuc, and Shakti Sinha. 2016. How to Get Them a Dream Job?: Entity-Aware Features for Personalized Job Search Ranking. In KDD. https://doi.org/10.1145/2939672.2939721

[16] Andrei Z. Broder, Adam Kirsch, Ravi Kumar, Michael Mitzenmacher, Eli Upfal, and Sergei Vassilvitskii. 2008. The Hiring Problem and Lake Wobegon Strategies. In SODA. https://doi.org/10.1137/07070629X

[17] Ravi Kumar, Silvio Lattanzi, Sergei Vassilvitskii, and Andrea Vattani. 2011. Hiring a Secretary from a Poset. In EC. https://doi.org/10.1145/1993574.1993582

[18] Esteban Arcaute and Sergei Vassilvitskii. 2009. Social Networks and Stable Matchings in the Job Market. In WINE. https://doi.org/10.1007/978-3-642-10841-9_21

[19] R. Rafter, K. Bradley, B. Smyth, "Automated Collaborative Filtering Applications for Online Recruitment Services," Adaptive Hypermedia and Adaptive Web-Based Systems, Lecture Notes in Computer Science, vol. 1892, pp. 363-368, 2000.

[20] R. Rafter, K. Bradley, B. Smyth, "Personalized Retrieval for Online Recruitment Services," Proceedings of the 22nd Annual Colloquium on Information Retrieval, IRSG 2000, Cambridge, UK, 5-7, Apr. 2000.

[21] Guo, Xingsheng, Jerbi, Houssem, O' Mahony, Michael P. : An Analysis Framework for Content-based Job Recommendation. 22nd International Conference on Case-Based Reasoning (ICCBR), Cork, Ireland, 29 September - 01 October 2014, 2014.

[22] Hong, W., Zheng, S., Wang, H.: Dynamic user profile-based job recommender system. Proceedings of the 8th International Conference Computer Science and Education, pp.1499–1503 (2013)

[23] Parul Aggarwal, Vishal Tomar, Aditya Kathuria: Comparing Content Based and Collaborative Filtering in Recommender Systems International Journal of New Technology and Research (IJNTR) ISSN:2454-4116, Volume-3, Issue-4, April 2017 Pages 65-67 [35] Su X, Khoshgoftaar T M. A Survey of collaborative filtering techniques. Advances in Artificial Intelligence, 2009, 421–425

International Journal of Computer Applications Technology and Research
Volume 7–Issue 06, 221-224, 2018, ISSN:-2319–8656

Information System Security Policy Studies as a Form of Company Privacy Protection

Rio Jumardi
Sekolah Tinggi Teknologi Bontang
Bontang, Indonesia

Abstract: Technology that interconnects computers in the world allows to be able to exchange information and data even communicate with each other in the form of images and video. The more valuable the information is required a security standard to maintain the information. Computer security target, among others, is as protection of information. The higher the security standards provided the higher the privacy protection of the information. Protection of employee privacy within a company is one factor that must be considered in the information systems implementation. Information system security policies include: System maintenance, risk handling, access rights settings and human resources, security and control of information assets, enterprise server security policy and password policy. The policies that have been reviewed, be a form of protection of corporate information.

Keywords: Computer Security, Information Assets, Information Systems, Privacy, Policy.

1. INTRODUCTION

Currently we are in the digital era where communication and information exchange takes place in a growing network of increasingly widespread. A Technology that interconnects computers in the world allows to be able to exchange information and data even communicate with each other in the form of images and video. The more valuable the information is required a security standard to maintain the information.

The system can be accessed with a current high availability is required, openness and distributed would have become a necessity for an integrated system. Information systems security management can reduce the occurrence of irregularities of the access rights by certain parties and misuse of data and information of an organization or company [1].

Computer security objectives are, among others the protection of information. The components of the security plan include: information security policies, standards and procedures, control of human resources management for information security, and control of information security technologies [2].

The higher the security standards provided the higher the privacy protection of the information. Protection of personal confidentiality of company employees is one of the factors that must be considered in the implementation of information systems. The making of information system security policy is expected to be a control of the organization's or company's behavior on the system.

Privacy is the ability of one or a group of individuals to retain life and personal affairs from the public, or to control the flow of information about themselves [3]. Privacy is sometimes associated with anonymity even though anonymity is especially appreciated by people known to the public. Privacy can be considered as an aspect of security.

Computer Crime is an unlawful act committed using a computer as a tool or computer as an object, whether to gain profit or not, to the detriment of the other party.

Computer crimes stipulated in the Act ITE provided for in Chapter VII of the act is prohibited. These deeds are categorized into several groups ie [4].

1. Unauthorized access
2. Unauthorized interception
3. Disturbance to computer data

Crimes that are closely linked to the use of technology based on these computers and telecommunications networks in several literatures and practices are grouped in several forms, including: [5]

1. Unauthorized Access Computer Systems and Service, Crimes committed by infiltrated into a computer network system illegally, without permission or without the knowledge of the owner of the computer network system he entered.
2. Contents, It is a crime by using data or information to the internet about a thing that is untrue, unethical, and may be considered unlawful or disturbing public order.
3. Forgery data, It is a crime to forge data on important documents stored as scriptless documents over the internet.
4. Cyber Espionage, is a crime that exploits the Internet network to conduct spying on other parties, by entering the target computer network system.
5. Cyber Sabotage and Extortion, Crime is done by making interference, destruction or destruction of a data, computer program or computer network system connected to the internet.
6. Offense Against Intellectual Property, This Crime is directed against intellectual property rights owned by others on the internet. For example, imitating the display on the web page of someone else's site illegally, broadcasting an information on the internet that turned out to be other people's trade secret information and so on.

The core of computer security is to protect computers and networks with the aim of securing the information in it. Computer security itself includes several aspects, among others: [6]

1. Authentication, the recipient of the information can ensure the authenticity of the message, that the message came from the person being asked for information. In other words, the information actually comes from the desired person.
2. Integrity, authenticity of messages sent over the network and it is certain that the information sent is not modified by unauthorized persons.

3. Non-repudiation, is related to the sender. The sender can not deny that it was he who sent the information.
4. Authority, the information residing on the network system can not be modified by unauthorized parties to access it.
5. Confidentiality, is an attempt to keep information from unauthorized persons. This secrecy usually relates to information provided to other parties.
6. Privacy, more towards personal data.
7. Availability, availability aspect relates to the availability of information when needed. Information systems that are attacked or uprooted can inhibit or eliminate access to information.
8. Access Control, this aspect relates to the way access to information is arranged. This is usually related to authentication and privacy issues. Access control is often done using a combination of user id and password or with other mechanisms.

Computer security provides requirements for computers that differ from most system requirements because they often take the form of restrictions on what computers should not do. This makes computer security is becoming more challenging because it is quite difficult to create a computer program to do everything what is designed to be done properly. Negative requirements are also difficult to meet and require in-depth testing for verification, which is impractical for most computer programs. Computer security provides a technical strategy for turning negative requirements into positive rules that can be enforced.

The main principle of information system security consists of confidentiality, integrity and availability or often called the CIA [7].

The general approach taken to improve computer security, among others, is to restrict physical access to the computer, implementing a mechanism on the hardware and operating system for computer security, as well as make your programming strategies to generate a reliable computer program.

ISO is one of the world's bodies that make standardization used by users or producers in a particular field. ISO 17799: 27002 is a standard that contains information security system settings[8].

Security clauses in ISO are [8]: Risk assessment and treatment, security policy, organization of information security, asset management, human resources security, physical and environmental security, communication and operation management, access control, information system acquisition, development and maintenance.

Information security aspects include the following ten aspects: security policy, security organization, classification and control assets, personnel security, physical and environmental security, communication and operations management, access control, system development and maintenance, business continuity management, harmonization.

2. RESEARCH METHODS

The research method used is a descriptive qualitative method that is the result of research presented in the form of narrative description. Qualitative approach done in this research is by detailing information security policy of the company information system with existing standards at ISO 17799:27002.

Data collection is done by direct observation in the field and direct interview with end users system and system manager in this case people who are competent in the field of information technology.

The Information Security Policy is defined as: An action plan for addressing information security issues, or a set of rules for maintaining certain information conditions or security levels [10].

Policy-making is based on a hierarchy of policies, standards, guidelines, procedures and practices.

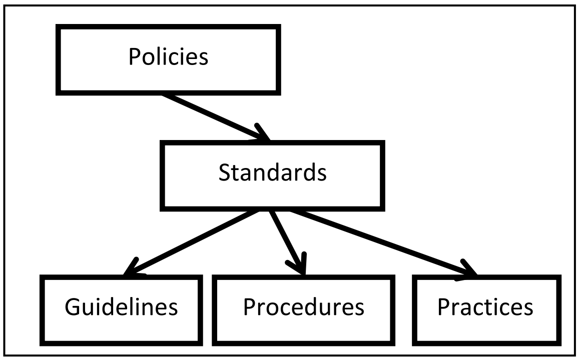


Figure 1. Policy-making Hierarchy [10]

The information security policy covers three general categories including Enterprise Information Security Policy (EISP), Issue Specific Security Policy (ISSP) and System Specific Policy (SSP) [10].

This research will discuss about EISP which includes system maintenance, risk handling, policy of access rights and human resources and security policy and control of information asset in the company and discuss about ISSP covering server security policy and password policy.

3. RESULTS AND DISCUSSION

From the introduction above, a study that discusses the information systems security policy will be one of the forms of privacy information company protection. Among the policies to be made are based on the ISO 17799: 27002 standard and also the standards issued by ID SIRTII include EISP, ISSP and SSP.

3.1 System Maintenance Policy

System maintenance policy is required to maximize maintenance of the running system, Company's system maintenance policy includes:

1. Objective: to ensure that the information system being implemented runs well.
2. Standard: used is the standard of ISO 17799: 27002 and Index KAMI as an evaluation tool.
3. Coverage: the implementation of this policy is directed to stakeholders and concerned employees in the Information Technology section as well as third parties who become vendors.
4. Guidelines for care: system maintenance should be in accordance with applicable guidelines.
5. Procedure: establish procedures related to system maintenance that include corrective care, adoptive care, preventive care and preventive care.
6. Monitoring: monitoring is required to monitor all activities related to the maintenance of the Company's systems.

3.2 Risk Management Policy

Risk management policies are required to address the risks that may arise during system implementation. The Company's risk management policy includes:

1. Objectives: identify and analyze possible risks that exist in the implementation of information systems in the company.
2. Standard: used is the standard of ISO 17799: 27002, ISO / IEC 27005, Octave Allegro Method.
3. Coverage: the implementation of this policy is intended for all employees in the corporate environment related to information assets.
4. Risk mitigation guidelines: risk handlers of the information systems and assets should be in compliance with applicable guidelines.
5. Procedures: establish procedures for risk management that include developing risk assessment criteria, developing an information asset profile, identifying containers from information assets, identifying problem areas, identifying threat scenarios, identifying risks, analyzing risks, and choosing a risk selection election approach.
6. Monitoring: monitoring is required to monitor all activities related to the Company's risk management.

3.3 Human Resource Policy

Human resource and access rights arrangements, policies are required to regulate the constraints of users of information systems within the Company. Human resource policies and corporate permissions settings include:

1. Objective: controlling user access of information systems by setting user permissions. Another purpose is to reduce risk for misuse of function or authority due to human error.
2. Standard: used is the standard of ISO 27002 and Information Technology Infrastructure Library (ITIL) V3.
3. Coverage: the implementation of this policy is directed to stakeholders and corporate leaders to determine or manage the determination of human resources by regulating access rights to the system.
4. Guidance: the determination of the right of access to the system shall be in accordance with the guidelines and rules applicable within the Company. Adjusted also with the ability of information systems to manage access rights.
5. Procedure: Establish procedures relating to the arrangement of access rights that include access requests, access grants, user identity monitoring, employee performance appraisal, employee job behavior, access restrictions, removal of access, access problems and access logging.
6. Monitoring: monitoring is required to monitor all activities related to human resource management and regulation of the access rights of information systems in the Company.

3.4 Security and Asset Control Policy

Security and asset control policies are required to manage the company's information assets. The company's information security and control policy include:

1. Objective: to provide protection for the company's assets based on the level of protection provided.
2. Standard: used is the standard of ISO 17799: 27002.
3. Coverage: the implementation of this policy is intended for stakeholders and corporate leaders and employees to the security of information assets in the use of information systems.
4. Guidelines: The guidelines for the security and control of information assets within a company's environment must be in accordance with the rules applicable to both the rules of the information system and the rules of the company.
5. Procedure: make procedures related to asset security and control of information assets include information classification and information responsibilities.
6. Monitoring: monitoring is required to monitor all activities related to the control of information assets in the Company.

3.5 Enterprise Server Security Policy

Another policy that must be considered by the company is the server security policy. This policy is necessary to maximize the security of data servers which will also directly maintain the confidentiality of Company data and employee privacy data against computer crimes that will harm the Company.

The Enterprise Server Security Policy includes:

1. Objective: maximize the security of the Company's information system from the server in use.
2. Standard: used is the standard of ISO 17799: 27002 and KAMI Index for evaluation tool.
3. Coverage: the implementation of this policy is directed to stakeholders and employees concerned in the Information Technology section
4. Guidance: The server configuration must be in accordance with applicable guidelines.
5. Procedure: make procedures related to server security that includes own server creation procedures, server storade procedures, server room security procedures, employees who served server room, and use of the server.
6. Monitoring: monitoring is required to monitor all activities related to the security of the Company's servers.

3.6 Password Policy

Password setting policy is required to set the password creation procedure and forget the password that is directly related to the security of the information system. The password setting policy includes:

1. Objective: maximize the security of corporate information systems through the use of secure passwords.
2. Standard: used is ISO 17799: 27002 standard.
3. Coverage: the implementation of this policy is directed to all company employees and system administrator sections
4. Guidance: Password settings must be in accordance with applicable guidelines.
5. Procedure: make procedures related to password setting that includes a procedure of making password, forgot password procedure, password reset procedure, captcha usage procedure, password active procedure and password combination rule.
6. Monitoring: monitoring is required to monitor all activities related to corporate information system's password setting activities.

4. CONCLUSION

Implementation of policies relating to the security of information systems is essential as a form of protection of corporate information.

The required policies include: System maintenance, risk handling, access control arrangement and human resources policies, security and control of information assets, server security policies and password policy. The protection provided not only for the Company's information, but also to the protection of the Company's personal privacy.

A Suggestion related to this research is a company can evaluate an information security policy using existing methods like KAMI Index, ITIL and other methods.

5. REFERENCES

[1] Wildan Radista Wicaksana, Anisah Herdiyanti , and Tony Dwi Susanto, "Pembuatan Standar Operasional Prosedur (SOP) Manajemen Akses Untuk Aplikasi E-Performance Bina Program Kota Surabaya Berdasarkan Kerangka Kerja ITIL V3 Dan ISO 27002," *Jurnal Sisfo*, vol. 06, no. 01, pp. 105-120, September 2016.

[2] Aan AlBOne, "Pembuatan Rencana Keamana Informasi Berdasarkan Analisis dan Mitigasi Risiko Teknologi Informasi," *Jurnal Informatika*, vol. 10, pp. 44-52, Mei 2009.

[3] "http://id.wikipedia.org/wiki/Kerahasiaan_pribadi," 2013.

[4] Ana Maria F. Pasaribu, "Kejahatan Siber Sebagai Dampak Negatif dari Perkembangan Teknologi dan Internet di Indonesia Berdasarkan Undang-undang No. 19 Tahun 2016 Perubahan atas Undang-undang No. 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik dan Perspektif Hukum Pidana," Universitas Sumatera Utara, Medan, Thesis 2017.

[5] Dodo Zaenal Abidin, "Kejahatan dalam Teknologi Informasi dan Komunikasi," *Jurnal Ilmiah Media Processor*, vol. 10, no. 2, pp. 509-516, Oktober 2015.

[6] Muhammad Siddik Hasibuan, "Keylogger Pada Aspek Keamanan Komputer," *Jurnal Teknovasi*, vol. 03, no. 1, pp. 8-15, 2016.

[7] Deni Ahmad Jakaria, R. Teduh Dirgahayu, and Hendrik, "Manajemen Risiko Sistem Informasi Akademik pada Perguruan Tinggi Menggunakan Metoda Octave Allegro," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, Yogyakarta, 2013, pp. E- 37-42.

[8] Deris Stiawan, "Kebijakan Sistem Informasi Manajemen Keamanan IT (Information Security Management Policy) Standard ISO 17799 : 27002," Universitas Sriwijaya, Palembang, 2009.

[9] Prof. Richardus Eko Indrajit, "ISO17799. Kerangka Standar Keamanan Infromasi," id-SIRTII, Jakarta, 2018.

[10] Iwan Sumantri. (2018, Mei) ID SIRTII. [Online]. http://cdn.woto.com/dsfile/49ba65ea-0804-46c3-aa21-86d156d167f9

On The Design of Web-Based Information and Booking System for Futsal Field Rental Business

Moechammad Sarosa
Department of Electrical Engineering
State Polytechnic of Malang
Malang, Indonesia

Verda Nurohmansah
Department of Electrical Engineering
State Polytechnic of Malang
Malang, Indonesia

Wahyu Indah Permana
Department of Electrical Engineering
State Polytechnic of Malang
Malang, Indonesia

Yoyok Heru Prasetyo Isnomo
Department of Electrical Engineering
State Polytechnic of Malang
Malang, Indonesia

Abstract: Limited free space for doing sports makes a lot of leased sports facilities, including futsal field. Nowadays, if someone wants to rent futsal field, a customer should make a reservation manually by going to the location or call the manager. This research tries to offer a solution in the form of web design for the owners of futsal field rental business. The designed website will provide registration services for field owners to promote their field to potential customers. The web will display information of several futsal fields that have registered in the form of location, schedule with its booking list and order procedure so that prospective customers could choose and place their order according to desired field and time. The web also features field location's map using google maps so that prospective buyer could know the exact location of each futsal field and the possible routes to the location. This information system has been tested on several computers and mobile phones connected to the internet. Testing result shows that this website has a responsive appearance. The site will adjust to the screen size when accessed devices with various specifications. The time delay in displaying a webpage depends on the quality of the network and the Android version of the device used.

Keywords: futsal field, information system, booking, service providers, rental business, reservation

1. INTRODUCTION

Limited available space to play sports makes futsal becomes a very popular sport since it does not need many spaces. This has led to a new business area such as futsal field rental which began to gain popularity. However, many of the field rental schedule management still be done manually and the booking still was done via phone or come directly to the location. With the rapid development of information technology, it is logical if an information system is developed to solve various problems in the futsal field rental [1] [2].

This research tries to develop a web application that provides services and booking information systems for managing the futsal field rental business. With this application, the futsal field entrepreneur (owner) can promote his futsal field, manage the booking and playing schedule [3].

The app also has supportive features, such as field-based searches, location map, rating the most popular field, and provide information and validation of each respective futsal, in which each owner has different rules. Another service provided by the application is information about the futsal field which is empty or has not been rented. However, the financial transactions made directly to related parties [4]

This application can help the field manager keep records such as field renter report and the customers can find information about the availability of the futsal field. Customers can also use Google Maps application which is included in the apps to show the location of registered places in the apps [5].

2. SYSTEM PLANNING

The web application is made using multiple software, such as Macromedia, Dreamweaver, Xampp, MySQL and also Google Maps API to show the locations of the futsal field [6] [7] [8].

This website divided into several parts, first is a user interface design that directly interacts with a user, including several commands to call the other part. Logic part is the other part of this website, in which it contains the core program and commands to access a database. While the database part that has some supportive add on an additional feature is called model part the three parts will be combined to form a webpage that satisfies user requirements [9].

This application is built according to System Development Life Cycle (SDLC) phases, in which an information system could support required application in the form of system design, building, and user presentation. SDLC has four main phases, namely: planning, analysis, design, and implementation [10]

2.1 System Description

This website is designed as a service media for renting futsal field aiming at providing web-based field rental service, including information media, rental facility, and rental schedule. The outline diagram of the application can be described as follows:

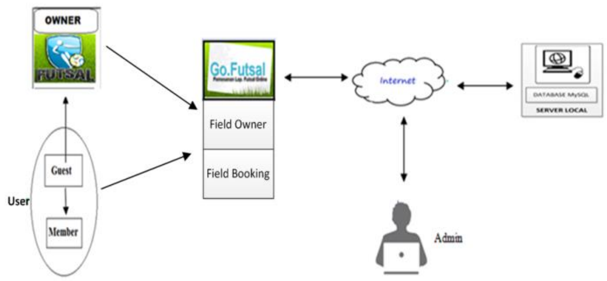


Figure 1. System block diagram

The system provides 4 types of access namely:

- Admin, in charge of managing the registration validation and viewing reports on the booking
- An owner, in charge of managing the booking schedule and the rental tariff.
- Guest will be able to register as a member, browse the information related to the futsal field registered on this website.
- Member will be able to book a field based on a futsal field search, date, price and time, and register his futsal field rental to this website.

Figure 2 shows the use case diagram which has 4 actors: Guest, Member, Owner, and Admin. Member, Owner, and Admin must login first to get into the system.

On the other hand, Guest can access the system without login. Guest can view the information about the futsal place, search field, schedule information, and see the location map. Guest can also register to become a member of the website.

Member can view the information about the futsal place, search field, schedule information, and see the location map. Member must login into the system to access the booking menu, see the history of booking and edit user accounts. The owner must login to access the menu which is changing the account data, add and change places data, add and change field data, add and change the data tariff, view the reports of the reservation, and validate the booking field.

Admin can change data account, validate the owner reservations, view report, view owner booking reports, and view member reports.

2.2 Use Case Diagram

Use Case diagram with 4 entities that cover admin, owner, user, and guest, shown in Figure 2.

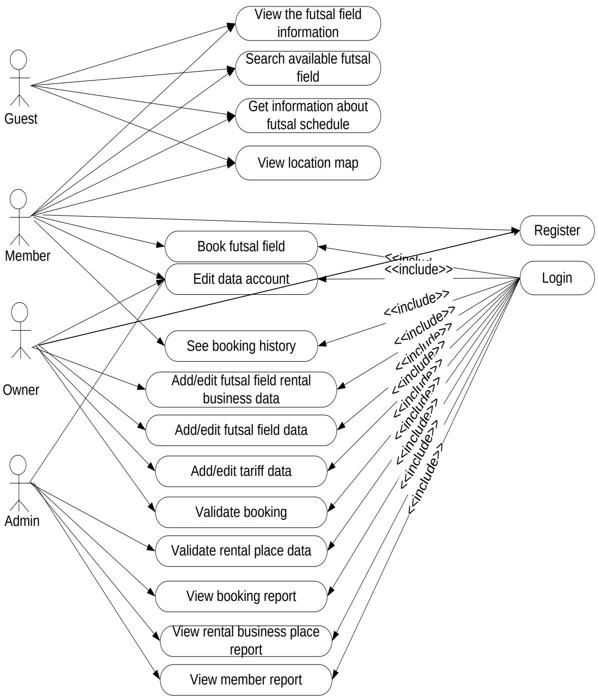

Guest
Member
Owner
Admin
View the futsal field information
Search available futsal field
Get information about futsal schedule
View location map
Register
Login
Book futsal field
Edit data account
See booking history
Add/edit futsal field rental business data
Add/edit futsal field data
Add/edit tariff data
Validate booking
Validate rental place data
View booking report
View rental business place report
View member report
<<include>>

Figure 2. Usecase diagram

2.3 System Flowchart

System flowchart built for user, owner, and admin entity shown in Figure 3, 4, and 5, respectively.

2.3.1 User

The futsal field booking application can be operated by the user namely a guest and members. When the application is running, the flow charts when users access the system shown in Figure 3.

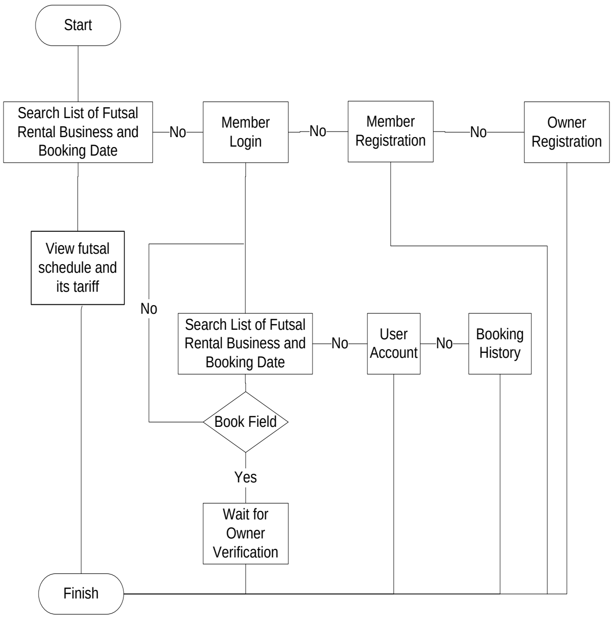

Start
Search List of Futsal Rental Business and Booking Date
No
Member Login
No
Member Registration
No
Owner Registration
View futsal schedule and its tariff
No
Search List of Futsal Rental Business and Booking Date
No
User Account
No
Booking History
Book Field
Yes
Wait for Owner Verification
Finish

Figure 3. Flowchart of "User"

The user stated in Figure 3 can be the guests and members. Guest can only view available futsal fields, and when a guest is interested, he can register as a member and can access the member menu such as booking the field, edit a user account, and view a history of booking. Guest can also register as an owner if he has a futsal field rental.

2.3.2 *Owner*

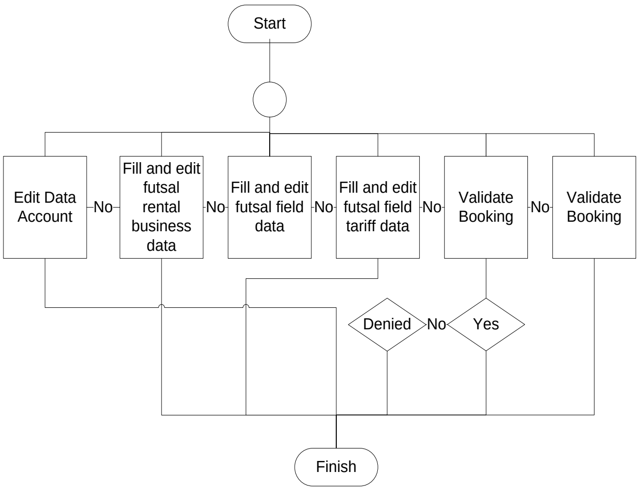

Start
Edit Data Account
No
Fill and edit futsal rental business data
No
Fill and edit futsal field data
No
Fill and edit futsal field tariff data
No
Validate Booking
No
Validate Booking
Denied
No
Yes
Finish

Figure 4. Flowchart of “Owner”

The owner can fill and update data account, fill the data and photos of the futsal field rental and the field itself, also fill and edit the futsal rental tariff, bevalidating field reservations booked by the user as valid or invalid, and also view the booking report.

2.3.3 *Admin*

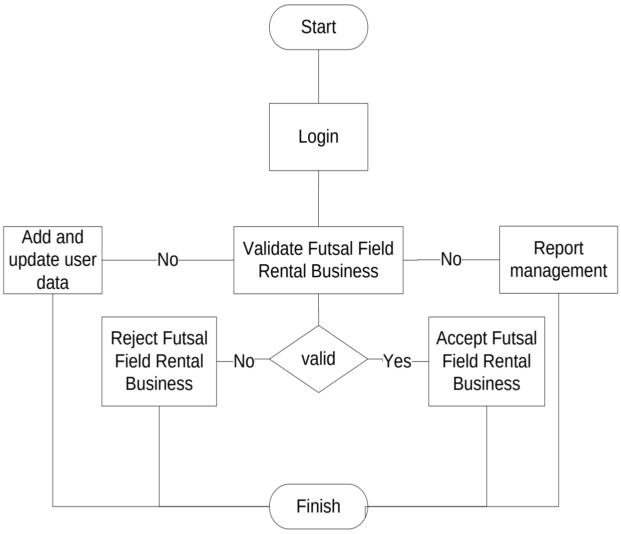

Start
Login
Add and update user data
No
Validate Futsal Field Rental Business
No
Report management
Reject Futsal Field Rental Business
No
valid
Yes
Accept Futsal Field Rental Business
Finish

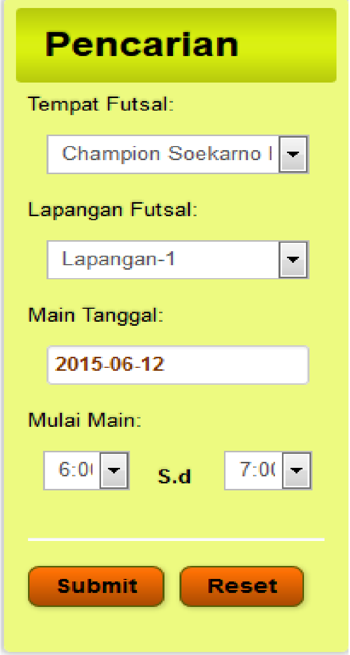
Figure 5. Flowchart of “Admin”

Admin should login with a particular username and password. An admin can add admin data and check the validation of the booking. If there is a new owner registered and the owner data is correct then admin will give “valid” status, so the futsal field belonging to the new owner can be displayed in a web. If the owner data is incorrect or does not fit the requirement, then admin will reject owner. Admin of report management division can view the booking report, rental places and members report.

3. TESTING OF APPLICATION MENU

Testing was done to observe the function of each menu, explained below:

3.1 Testing Booking Futsal Field Menu



Pencarian

Tempat Futsal:

Champion Soekarno I

Lapangan Futsal:

Lapangan-1

Main Tanggal:

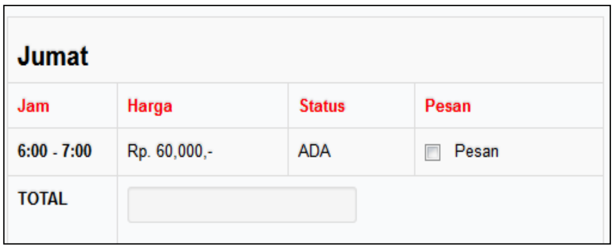
2015-06-12

Mulai Main:

6:0(S.d 7:0(

Figure 5. Search futsal field rental business

This page showedall the registered futsal field rental place which is available to be booked. If a user wants to book the futsal field, a user should select the futsal field rental place, desired date and time a user starts to play. For example, book the Champion SoekarnoHatta futsal field at June 12th, 2015, 06:00-07:00 a.m.



Jumat

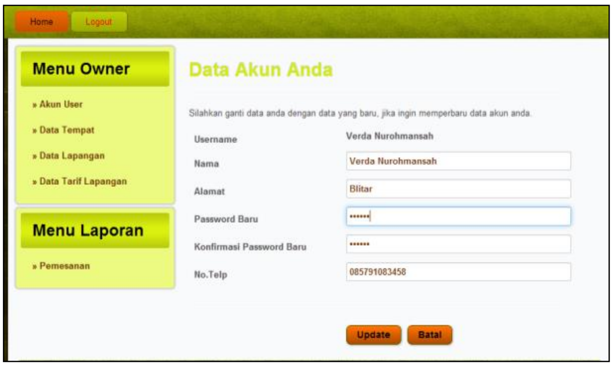
Jam	Harga	Status	Pesan
6:00 - 7:00	Rp. 60,000,-	ADA	□ Pesan
TOTAL			

Figure 6. Booking futsal field

On the futsal field schedule page, if the selected field is available then a user could choose “book” and the total price user should pay appears automatically according to the booking tariff.

3.2 Testing Registration of Futsal Field Rental Business

For an owner who wants to register their futsal field rental business, the owner should first enter the registration page of a new owner and fill in the data to be able to log into the web application. After login into the web, an owner can fill all field data futsal namely user account data, place, fields and field tariff data.



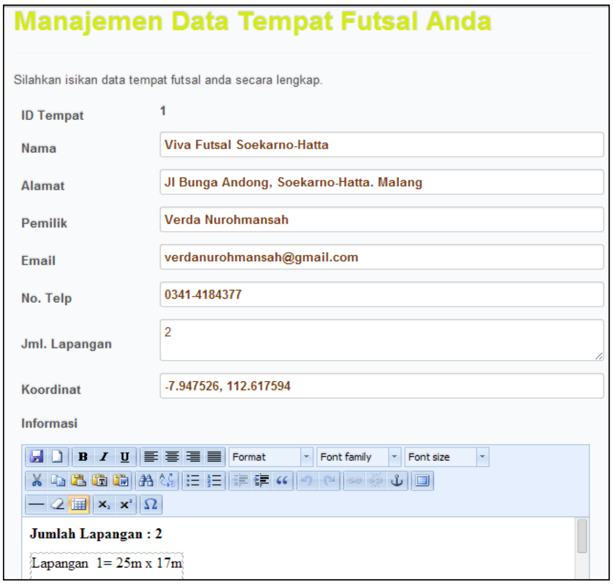
Data Akun Anda

Silahkan ganti data anda dengan data yang baru, jika ingin memperbaru data akun anda.

Username	Verda Nurohmansah
Nama	Verda Nurohmansah
Alamat	Blitar
Password Baru	••••••
Konfirmasi Password Baru	••••••
No.Telp	085791083458

Figure 7. Menu owner's data account

If an owner wants to change the new data then he must change the desired data and press the update button to save the data, so that the data can be updated.



Manajemen Data Tempat Futsal Anda

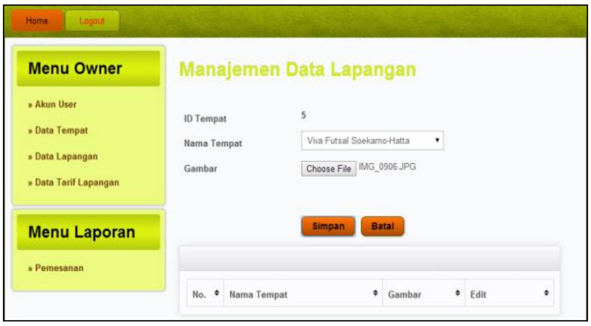
Silahkan isikan data tempat futsal anda secara lengkap.

ID Tempat	1
Nama	Viva Futsal Soekarno-Hatta
Alamat	Jl Bunga Andong, Soekarno-Hatta. Malang
Pemilik	Verda Nurohmansah
Email	verdanurohmansah@gmail.com
No. Telp	0341-4184377
Jml. Lapangan	2
Koordinat	-7.947526, 112.617594
Informasi	

Jumlah Lapangan : 2

Lapangan 1= 25m x 17m

Figure 8. Rental business' data menu



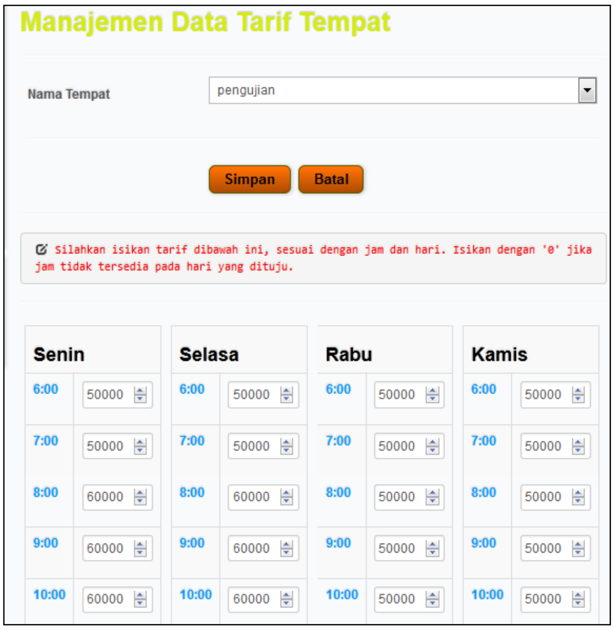
Manajemen Data Lapangan

ID Tempat	5
Nama Tempat	Viva Futsal Soekarno-Hatta
Gambar	IMG_0906.JPG

No.	Nama Tempat	Gambar	Edit

Figure 9. Futsal field' data menu

An owner who has successfully fill in the data about his futsal field could continue to the field data menu in which he will upload the futsal filed' photos.



Manajemen Data Tarif Tempat

Nama Tempat: pengujian

Silahkan isikan tarif dibawah ini, sesuai dengan jam dan hari. Isikan dengan '0' jika jam tidak tersedia pada hari yang dituju.

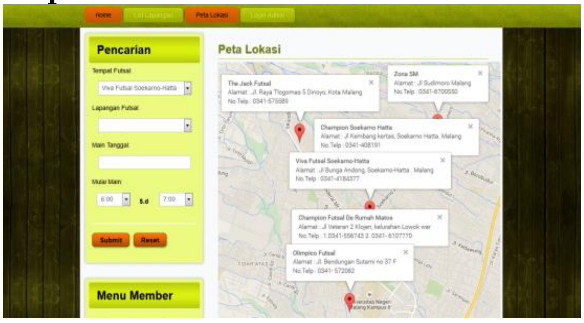
Senin		Selasa		Rabu		Kamis	
6:00	50000	6:00	50000	6:00	50000	6:00	50000
7:00	50000	7:00	50000	7:00	50000	7:00	50000
8:00	60000	8:00	60000	8:00	50000	8:00	50000
9:00	60000	9:00	60000	9:00	50000	9:00	50000
10:00	60000	10:00	60000	10:00	50000	10:00	50000

Figure 10. Futsal field tariff menu

3.3 Web Design Responsivity Testing

Responsivity testing is aimed to obtain the information system display size suitability on a variety of different devices such as computers and smartphones. The information system can adjust the display size on the device screen used which influence the displayed information.

3.4 The Test result for screen size >1280 pixels



Peta Lokasi

The Jack Futsal
Alamat : Jl. Raya Tlogomas 5 Dinoyo, Kota Malang
No Telp : 0341-575589

Zona SM
Alamat : Jl Sudimoro Malang
No Telp : 0341-6700550

Champion Soekarno Hatta
Alamat : Jl Kembang kertas, Soekarno Hatta. Malang
No Telp : 0341-408191

Viva Futsal Soekarno-Hatta
Alamat : Jl Bunga Andong, Soekarno-Hatta . Malang
No Telp : 0341-4184377

Champion Futsal De Rumah Matos
Alamat : Jl Veteran 2 Klojen, kelurahan Lowok war
No Telp : 1.0341-556743 2. 0341-610777D

Olimpico Futsal
Alamat : Jl. Bendungan Sutami no 87 F
No Telp : 0341- 572062

Tempat Futsal	Viva Futsal Soekarno-Hatta
Mulai Main	6:00 s.d 7:00

Figure 11. Screen size display in PC

Figure 11 shows the display in which the screen size is larger than 1280 pixels or computer screen. At this size, the map could show all registered futsal place in Malang, such as The Jack Futsal, Zone SM, Champion Soekarno-Hatta, Viva Futsal Champion de house Malang Town Square and Futsal Olimpico.

3.5 The Test result for screen size < 768 pixels

Figure 12 shows the display for screen size <768 pixels or screen for tablets. At this size, a map can only display 4 out of 6 (67%) registered futsal place in the city of Malang, which are The Jack Futsal, Zone SM, Champion SoekarnoHatta and Viva Futsal.

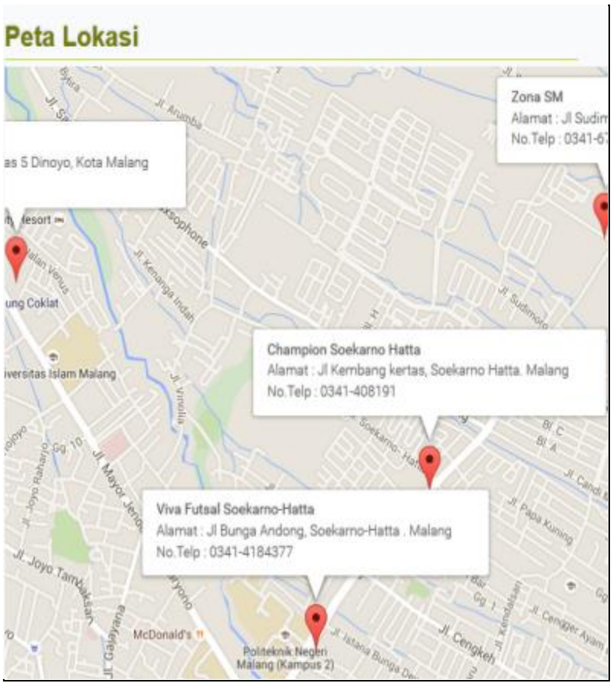

Peta Lokasi
Zona SM
Alamat : Jl Sudim
No.Telp : 0341-67
as 5 Dinoyo, Kota Malang
Champion Soekarno Hatta
Alamat : Jl Kembang kertas, Soekarno Hatta. Malang
No.Telp : 0341-408191
Viva Futsal Soekarno-Hatta
Alamat : Jl Bunga Andong, Soekarno-Hatta . Malang
No.Telp : 0341-4184377
McDonald's
Politeknik Negeri Malang (Kampus 2)

Figure 12. Screen size display in tablet

3.6 The Test result to screen size < 480 pixels

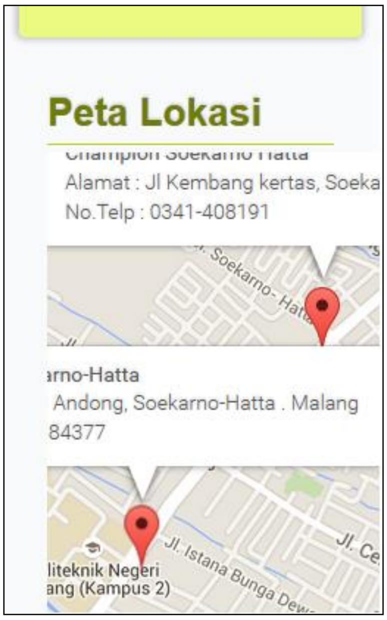


Figure 13. Screen size display in a smartphone

Figure 13 shows the display for screen size <480 pixels or screens for smartphones. On this size, a map can only display 2 area out of 6 (33%) registered futsal rental place in the city of Malang which are Viva Champion Futsal Futsal and SoekarnoHatta.

3.7 Testing on a variety of device

Testing was conducted on some smartphones at the same time to determine the feasibility of the new application. Testing was conducted using 5 different smartphone versions and brands, types of Internet data packets are the same at the same time and in the same place.

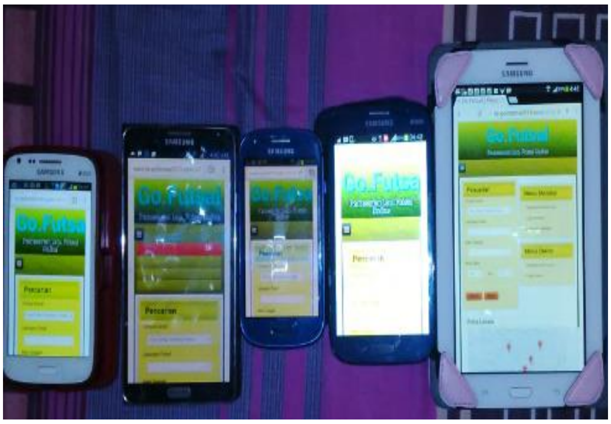


Figure 14. Testing on a variety of device

Table 1. Delay In Android Version 4.2.3

	Hardware Specification (RAM and Processor)	Android Version	Downlink Rate	Delay (s) Thursday (25/06/2015) – Sunday (28/06/2015) 3.44 p.m				Average Delay (s)
				Thu	Fri	Sat	Sun	
1	Smartphone A	4.2.3	Up to 7.2 Mbps	3	3	4	3	3.25
2	Smartphone E	4.2.3	Up to 7.2 Mbps	3	3	5	4	3.75
Total Average Delay (s)								3.5

From the above results, network traffic and smartphone brands during rush hour affect the application delay. In HSDPA+ network which its downlink rate is up to 7.2 Mbps, there is a delay on the Smartphone A of 3.25 seconds and in Smartphone E with a delay of 3.75 seconds, so that the average delay is 3.5 seconds.

From Table 2, network traffic and smartphone brands during rush hour also affect the application delay in older android version. In HSDPA network which its downlink rate is up to 7.2 Mbps, there is a delay on the Smartphone C of 5.25 seconds and in Smartphone B with a delay of 7 seconds, so that the average delay is 6.25 seconds.

Table 2. Delay In Android Version 4.2.2

No	Hardware Specification (RAM and Processor)	Android Version	Downlink Rate	Delay (s) Thursday (25/06/2015) –Sunday (28/06/2015) 3.44 p.m				Average Delay (s)
				Thu	Fri	Sat	Sun	
1	Smartphone B	4.2.2	Up to 7.2 Mbps	6	6	10	6	7
2	Smartphone C	4.2.2	Up to 7.2 Mbps	5	6	5	5	5.25
3	Smartphone D	4.2.2	Up to 7.2 Mbps	6	5	9	6	6.5
Total Average Delay (s)								6.25

4. CONCLUSION

The applications could be used properly in accordance with the design and there is no mistake in the process of displaying information. Based on test results performed on the screen size> 1280 pixels, <768 and <480 pixels, the system had a responsive view, means that the displayed information can adjust to the screen size when accessed on a variety of devices such as computers and smartphones. The percentage of displayed map image by the device on screen size> 1280 pixels is 100%, <768 pixel is 67% and screen size <480 pixel is 33%. From the smartphone-based test obtained in the same time with the same brand and android version of 4.2.3, the average delay time is 3.5 seconds, while in android version 4.2.2, then average delay time reaches 6.25 seconds. It can be concluded that the slow and fast access time depends on the android version of the smartphone.

In the future, the system can be developed with the facility of payment via e-banking. and can be equipped with socialmedia to remind the user who booked the field.

5. BIBLIOGRAPHY

[1] Zakaria R.Z., "SistemInformasiPenyewaanLapangan Futsal Berbasis Web Dan SMS Gateway (StudiKasus Goal Arena Futsal).," *JurnalUniversitas Pembangunan Nasional Veteran JawaTimur.*, 2013.

[2] Kristiandi H., "Pembangunan Aplikasi Mobile PencarianPersewaanLapangan Futsal di Yogyakarta BerbasisLokasi," UniversitasAtma Jaya, Yogyakarta, Doctoral Dissertation 2014.

[3] Ruhmawan, A.R. and Nurwidyantoro, A., "SistemInformasiPemesananLapangan Futsal Berbasis Web," UniversitasGadjahMada, Yogyakarta, Doctoral Dissertation 2015.

[4] Maimunah , Hariyansyah , and Jihadi, G, "RancangBangunSistemAplikasiPenyewaanLapangan Futsal Berbasis Web," in *Seminar NasionalTeknologiInformasidan Multimedia, STMIK Amikom*, Yogyakarta, 2017, pp. 4.7-7 - 4.7-12.

[5] Jannah, E.N. and Hidayah, A., "SistemTerintegrasiBerbasis Web untukPencariandanPemesananKelompokSeniPertunjukan.," *JurnalNasionalTeknikElektrodanTeknologiInformasi (JNTETI)*, vol. 5(4)., 2016.

[6] Hu, S. and Dai, T., "Online Map Application Development Using Google Maps API, SQL Database, and ASP.NET.," *International Journal of Information and Communication Technology Research*, vol. 3(3)., 2013.

[7] Anhar ST., *PanduanMenguasai PHP dan MySQL secaraotodidak*. Jakarta, Indonesia: Mediakita, 2010.

[8] Nugroho T., *Practice Guide PHP on Windows.* Jakarta, Indonesia: PT. Elex Media Komputindo, 2009.

[9] Zamahsari, Sarosa M., and Nurwasito H., "Concept of Designing an Optimized Pull Model View," *International Journal of Computer Applications (IJCA)*, vol. 57(20), pp. 9-13, 2012.

[10] Sommerville I, *Software Engineering 9thEdition.*: China Machine Press, 2011.

International Journal of Computer Applications Technology and Research
Volume 7–Issue 06, 231-240, 2018, ISSN:-2319–8656

Multiple Use of Surface Water Resources and B
Colonization of Water Bodies - Case (II) Ariam R
Other Tributaries in Ezinihite Mbaise

Ibezue Victoria C (Ph.D)
Department of Geology, Faculty of
Physical Sciences,
COO, University Uli, P.M.B. 02,
Uli, Nigeria

Ndukwe John O
Department of URP, Faculty of
Environmental Sciences, COO,
University Uli, P.M.B. 02, Uli,
Nigeria

Nwabineli E
[2]Department of C
Akanu Ibiam F
Unwana, Ebon

Abstract

Water samples collected along the water courses of surface water sources of domestic water supply in *Ezinihite M*
bacterial species inventory and total viable count (TVC) using the multiple test tube technique and colony coun
covered include Ariam River and other tributaries that constitute the bulk of surface water resources in the area. Eight spe
E-coli, staphylococcus aureus, salmonella, and fecal streptococci among others were identified. Total viable counts gav
when compared o the standards as set by the world health organization (WHO). The microbial population explosion in t
the multiple activities within and around the river also the uses including wash off from abattoirs carrying abattoir wast
domestic wastes dumped along the recharge path, others include in stream fermentation of food stuff and general laundry
automobiles. All these make sufficiently available to enhance microbial growth. Surface water use should be monitored
and proper management of watershed will control this trend of colonization of public water supply sources and in tu
water borne infections.

Keywords: tributaries, bacteria colonization, species inventory, total viable counts, surface water and fermentat

Introduction

The colonization of water by organisms depends on the physical and chemical state of the water [2]. Disease causing organisms enter water supplies via human wastes and sewage [5]. One of the major challenges facing environmental managers, hydrologists, water resource analysts and the allied professionals in Nigeria today is the problem of surface water pollution. And rapid urbanization, domestic and industrial activities constitute the sources of pollutants to urban and rural rivers. Surface water supplies vary in quality relative to the seasons, climatic conditions and uses [3], [6] and [4]. The area of study is drained by a major river - Ariam and three streams as its tributaries. These surface water sources are subjected to multiple uses without monitoring, as though they constitute a center for all human activities in the community; from what may be termed as general laundry (for motor bikes, fermented starches, clothes etc), through agricultural food processing, domestic uses, recreational for children, gravel and sand mining, navigation, fishing and it plays host to waste dumps including wash offs from agricultural lands. This wide range of unmonitored uses places the quality of the surface waters in the area in doubt, hence this work tried to estimate the bacteria load in the water supply sources in Ezinihite Mbaise as to advice on management options.

Procedure for Data Collection

In accord with the standard procedure fo
sampling of American Public Health Asso
were collected from ten (10) locations acro
course. The samples were subjected to serie
from serial dilution, plating, incubation, ino

Incubation

It is done to test for species inventory, t
incubated in Mac Conkey broth medium at 3
using the multi test-tube technique. The test
nine species of bacteria including E-coli,
salmonella and shigella, yeast and mo
Klebsiella pneumonia, pseudomonas a
streptococci. Bacteria was presumed prese
inverted tubes indicating the presence of E-
form spp on Petri dishes, acid pH change
Streptococcus Bacilli and translucent aga
Typhi. Aerobic incubation in Salmonella Ag
showed colourless and translucent appearanc
nor produce H_2S which is indicative of Shig
species colonies. Translucent with a black c

indicates the presence of Proteus Mirabilis and most Salmonella spp. Incubation in Cled Agar: yellow opaque colonies indicates presence of E-coli; extremely mucoid colonies varying in colour from yellow to whitish blue indicates Klebsiella spp; yellow to green colonies indicates Pseudomonas Aeuriginosa; yellow colonies indicates the presence of streptococci faecalis whereas deep yellow colonies presents staphylococcus aureus spp.

Relevance of the Study

This work is recommended for both scholars in water resources and environmental pollution. The innocence with which the users carry out their daily activities within and around water bodies tells a lot about their level of ignorance of the severity of the effects of the said activities. Coupled with the fact that, they are victims of their own doing as it is them that use the water for drinking and other domestic purposes. The various uses predisposes them to myriads of diseases from the organisms identified. Therefore, this paper if made available to public by publishing, will serve to educate the members of the public of the implication of wrong approaches to public utilities. To think that water consumption accounts for many diseases depending on the source.

Biomass Estimation

The heterotrophic plate count was used to d
count of the individual bacterial species iden
2ml of water sample was adapted in 10ml of
on flat dish at 45^{o}C. The culture was allow
plates inoculated at different temperature a
22^{o}Cfor 72hours and the other at 37^{o}C for 24
appeared were counted on the microscope a
Total Viable Counts for the individual water

Membrane Filtration

Further step was taken as to confirm the pres
species of coli-form bacteria, so that the w
through the membrane filtration equipment a
Lauryl Sulphate Agar plates aerobically for 2
for 14hours at 44^{o}C, a yellow colony con
Positive methyl red reaction (indicative of a r
and negative reaction/non citrate utilization
aid confirmatory tests for coli-form.

Findings

The table below shows that about eight
bacteria were present in varying number of
which are collected from different locations a

Table 1 below reflects the result of the species inventory and the total viable counts (TVC) for the individual species identified.

Table 1: **Bacteria species inventory and the total viable counts (TVC)**

SAMPLES	STREPTOCOCCI (FECAL)	PROTEUS MIRABILIS	E COLI	KLEBSIELLA PNEUM	SALMONELLA SHIGELLA	CLOSTRIDIUM	PSEUDO AERUINOSA
I	6.3 X 10^7	5.2 X 10^7	8.1 X 10^7	2.4X 10^7	5.0X 10^7	1.2 X 10^7	6.4X 10^7
Ii	7.4 X 10^7	5.0 X 10^7	6.5 X 10^7	2.9 X 10^7	2.7 X 10^7	1.4 X 10^7	5.3 X 10^7
Iii	8.3 X 10^7	4.1 X 10^7	8.4 X 10^7	6.6 X 10^7	3.5 X 10^7	2.6 X 10^7	5.1 X 10^7
Iv	5.4 X 10^7	4.5 X 10^7	6.5 X 10^7	4.3 X 10^7	4.7 X 10^7	2.3 X 10^7	4.5 X 10^7
V	4.8 X 10^7	3.2 X 10^7	5.5 X 10^7	4.7 X 10^7	5.5 X 10^7	3.5 X 10^7	4.9 X 10^7
Vi	7.0 X 10^7	3.6 X 10^7	6.3 X 10^7	5.1 X 10^7	7.2 X 10^7	4.5 X 10^7	5.0 X 10^7
Vii	7.2 X 10^7	4.4 X 10^7	6.2 X 10^7	6.0 X 10^7	6.3 X 10^7	3.0 X 10^7	5.4 X 10^7
Viii	6.7 X 10^7	4.5 X 10^7	5.1 X 10^7	6.1 X 10^7	4.1 X 10^7	2.7 X 10^7	5.7 X 10^7
Ix	5.4 X 10^7	5.0 X 10^7	6.0 X 10^7	6.5 X 10^7	7.7 X 10^7	4.8 X 10^7	4.3 X 10^7
x	6.2 X 10^7	6.2 X 10^7	7.4 X 10^7	5.5 X 10^7	5.8 X 10^7	3.3 X 10^7	6.2 X 10^7

Whereas figure 1 below shows the extent of colonization of the surface water sources by the bacteria species in terms courses.

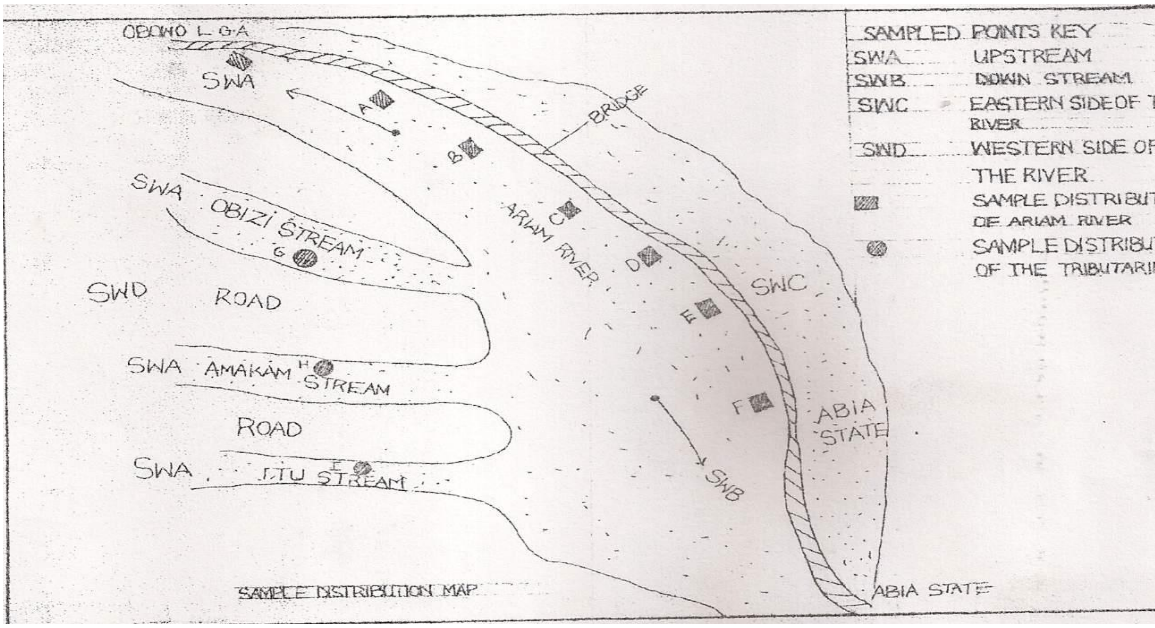


Figure 1. Map of the study area showing sample locations

Figures 2, 3 and 4 below are a graphical representation of some of the data in table which tries to compare TVC within and across samples. Observation shows that there are variations in TVC between and across samples and from location to location as well as between water bodies. The variation fron attributed to the proximity to the variou organisms, as regards the points of entran from human activities going in those place

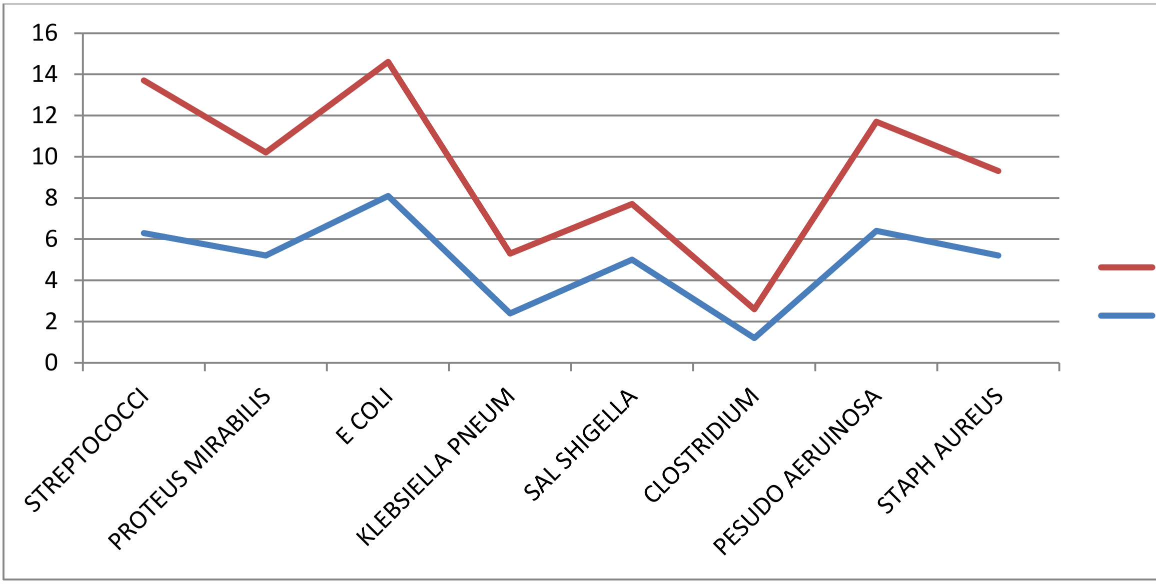


Figure 2. Graphical representation comparing results from two locations (A&B) along river Ariam water course

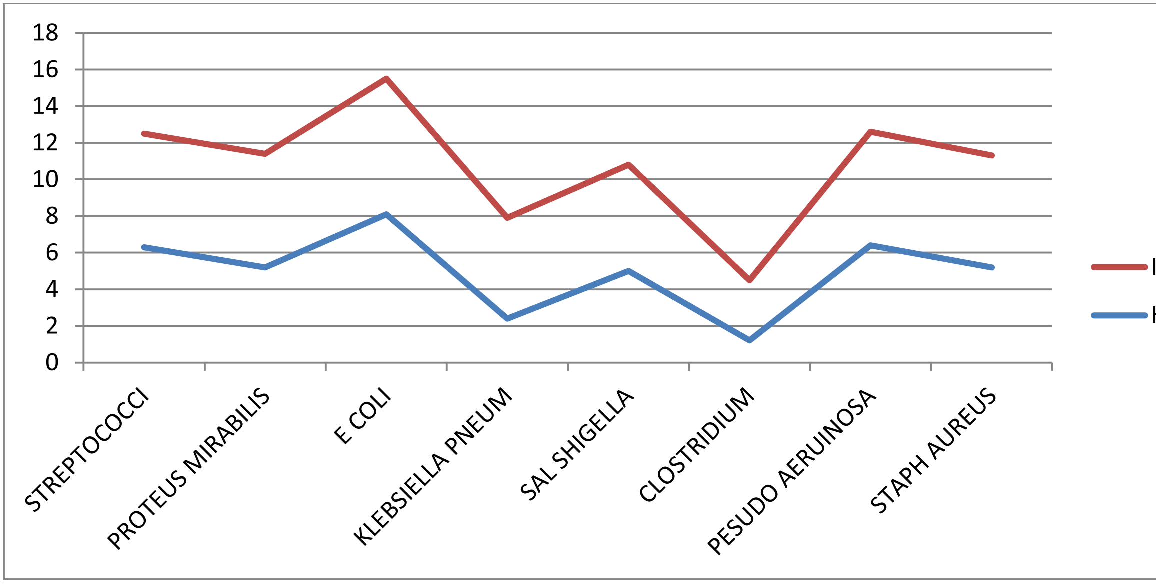


Figure 3. Graphical representation comparing results from two locations I(Itu) and H (Amakam) str

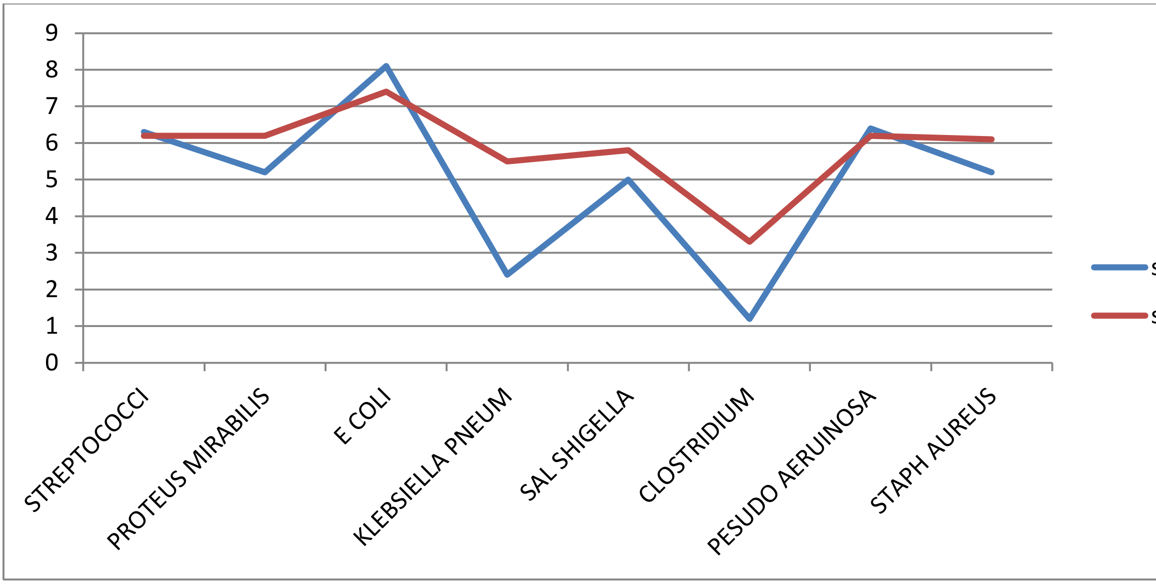


Figure 4. Graphical representation comparing results from two locations Itu stream and Ariam river

Figure 5 below is a plot of the biological water quality indicator (E-coli) and the fecal streptococci against other species
that the two are at all times and all places high in TVC indicating strong relationship between its occurrence and human

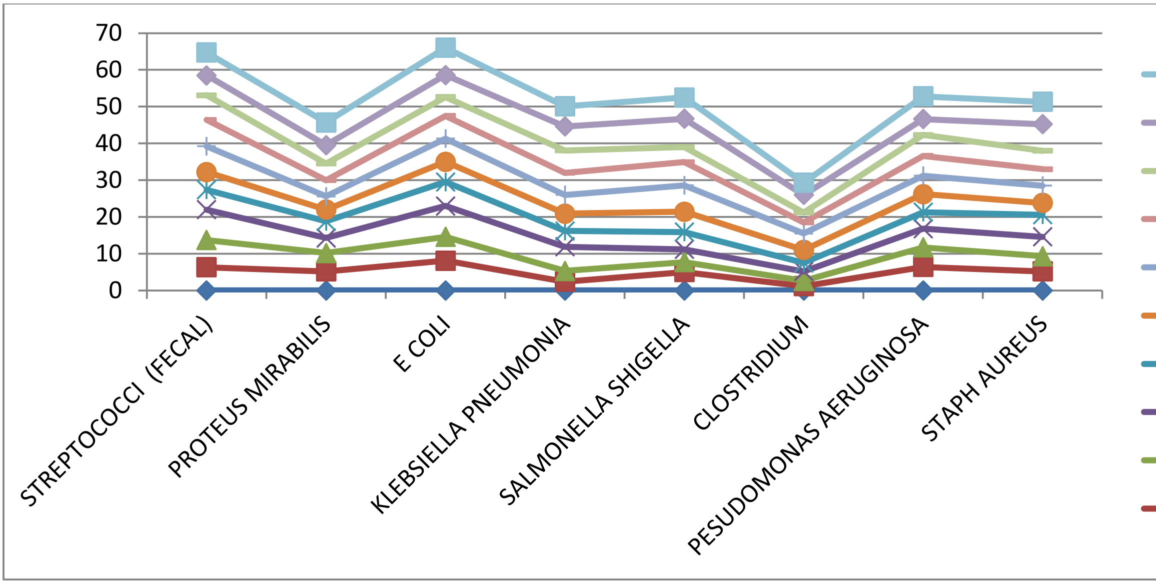


Figure 5. Graphical representation comparing TVC of individual species at various locations

Implication of Findings

The extent of colonization in terms of spread and the numerical strength of the organisms shows that the water supply source in the study area is unfit for drinking and most domestic uses as it is also evident that using such water in washing of food stuff, bathing predisposes the very user to infection. The water use pattern designate in the study area serve as a recycling process to these organisms concerning the severity of routine uses whic

from drinking, through other domestic uses

wastes (directly discharged by persons int

through runoff) and wastes (domestic, indus

Summary

The surface water sources of water supply in the study area require massive disinfection along the length and breadth of the water bodies. And the direct use such as washing fermented starches and other biodegradables as well as direct discharge of wastes by persons of human wastes into the surface water bodies or on the recharge area should be discouraged by community leaders. Dislodging industrial wastes into the surface water sources should be disallowed while protecting the recharge paths from agricultural land wash offs and the water bodies from any harmful activity. The surface water sources can be kept from direct use especially the upstream side and water can be pumped out and piped away from source to various locations and as the need arises. Even canals/impoundment pits can be used to extract water from the main rivers/streams to provide for other uses.

References

[1] APHA-AWWA-WPCF (1995): Stan examination of water and waste waters, 1 Health Association Washington DC.

[2] Okechukwu,G. C. (1983): "the effect hydrology and water resources

[3] Onwuekwe, (2004): " bacteriologica characteristics of some
selected points along the River Niger" B.SC Department of Production Technology, Nna Awka

[4] Nwackukwu, Okereke and Ukpabi (19 Otamiri River, Environmental Review, Vol.

[5] Wagner, E. G. and Lanoix, J.N. (1959): areas and small
communities", World Health Organization,

[6] Watts,J. (1953): Availability of water an Journal of Public Health, 43:728

Semantic Similarity Measures between Terms in the Biomedical Domain within frame work Unified Medical Language System (UMLS)

Abdelhakeem M. B. Abdelrahman
Sudan University of Science and Technology
Collage of Graduate Studies Khartoum, Sudan

Dr. Ahmad Kayed
Department of Computing and Information Technology
Sohar University, Sohar, Oman

Abstract

The techniques and tests are tools used to define how measure the goodness of ontology or its resources. The similarity between biomedical classes/concepts is an important task for the biomedical information extraction and knowledge discovery. However, most of the semantic similarity techniques can be adopted to be used in the biomedical domain (UMLS). Many experiments have been conducted to check the applicability of these measures. In this paper, we investigate to measure semantic similarity between two terms within single ontology or multiple ontologies in ICD-10 “V1.0” as primary source, and compare my results to human experts score by correlation coefficient.

Keywords: Information extraction, biomedical domain, semantic similarity techniques, Unified Medical Language System (UMLS), and Semantic Information Retrieval (SIR).

1. INTRODUCTION

Ontology is test bed of semantic web, capturing knowledge about certain area via providing relevant concept and relation between them. Quality metrics are essential to evaluate the quality. Metrics are based on structure and semantic level. At the present the ontology evaluation is based only on structural metrics, which has not been very appropriate in providing desired results.

Semantic similarity measures are widely used in Natural Language Processing. We show how six existing domain-independent measures can be adapted to the biomedical domain. Semantic similarity techniques are becoming important components in most intelligent knowledge-based and Semantic Information Retrieval (SIR) systems [1]. Measures and tests are provided to define how we can measure the "goodness" of ontology or its resources. Many experiments have been conducted to check the applicability of these measures [4].

General English ontology based structure similarity measures can be adopted to be used into the biomedical domain within UMLS. New approach for measuring semantic similarity between biomedical concepts using multiple ontologies is proposed by Al-Mubaid and Nguyen [2, 3]. They proposed new ontology structure based technique for measuring semantic similarity between single ontology and multiple ontologies in the biomedical domain within the frame work of Unified Medical Subject Language System (UMLS). Their proposed measure based on three features [2]: first Cross modified path length between two concepts. Second, new features of common specificity of concepts in the ontology. Third Local ontology granularity of ontology cluster.

2. BIOMEDICAL DOMAIN ONTOLOGIES

Most of the semantic similarity techniques work in the biomedical domain uses only ontology (e.g. MeSH, SOMED-CT) for computing the similarity between the biomedical terms[9]. However, in this work we use ICD- 10 ontology as primary source to computing the similarity between concepts in biomedical domain.

International Classification of Diseases (ICD): The newest edition (ICD- 10) is divided into 22 chapters: (Infections, Neoplasm, Blood Diseases, Endocrine Diseases, etc.), and denote about 14,000 classes of diseases and related problems. The first character of the ICD code is a letter, and each letter is associated with a particular chapter, except for the letter D, which is used in both Chapter II, Neoplasm, and Chapter III, Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism, and the letter H, which is used in both Chapter VII, Diseases of the eye and adnexa and Chapter VIII, Diseases of the ear and mastoid process. Four chapters (Chapters I, II, XIX and XX) use more than one letter in the first position of their codes. Each chapter contains sufficient three-character categories to cover its content; not all available codes are used, allowing space for future revision and expansion. Chapters I–XVII relate to diseases and other morbid conditions, and Chapter XIX to injuries, poisoning and certain other consequences of external causes. The remaining chapters complete the range of subject matter nowadays included in diagnostic data. Chapter XVIII covers Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified. Chapter XX, External causes of morbidity and mortality, was traditionally used to classify causes of injury and poisoning, but, since the Ninth Revision, has also provided for any recorded external cause of diseases and other morbid conditions. Finally, Chapter XXI, Factors influencing health status and contact with health services, is intended for the classification of data explaining the reason for contact with health-care services of a person not currently sick, or the circumstances in which the patient is receiving care at that particular time or otherwise having some bearing on that person’s care [8, 10].

3. SEMANTIC SIMILARITY TECHNIQUES CHALLENGES IN THE BIOMEDICAL DOMAIN

Most of existing semantic similarity techniques that used ontology structure as the primary source can’t measure the similarity between terms using single ontology or multiple ontologies in the biomedical domain within frame work Unified Medical Language System (UMLS). However, some of the semantic similarity techniques have been adopted to biomedical domain by incorporating domain information extracted from clinical data or medical ontologies.

4. RELATED WORK

4.1 Rada et al. Proposed semantic distance as a potential measure for semantic similarity between two concepts in MeSH, and implemented the shortest path length measure, called CDist, based on the shortest distance between two concept nodes in the ontology. They evaluated CDist on UMLS Metathesaurus (MeSH, SNOMED, ICD9), and then compared the CDist similarity scores to human expert scores by correlation coefficients.

4.2 Caviedes and cimino. [11] Implemented shortest path based measure, called CDist, based on the shortest distance between two concepts nodes in the ontology. They evaluated CDist on UMLS Metathesaurus (MeSH, SNOMED, ICD9), and then compared the CDist similarity scores to human expert scores by correlation coefficient.

4.3 Pedersen et al.[1] Proposed semantic similarity and relatedness in the biomedicine domain, by applied a corpus-based context vector approach to measure similarity between concepts in

SNOMED-CT. Their context vector approach is ontology-free but requires training text, for which, they used text data from Mayo Clinic corpus of medical notes.

4.4 Wu and Palmer Similarity Measure [11] proposed a new method which define the semantic similarity techniques between concepts C_1 and C_2 as

$$\mathrm{Sim}\,(C_1, C_2) = 2 \times \frac{N_3}{N_1 + N_2 + 2 \times N_3} \qquad (1)$$

Where

N_1 is the length given as the number of nodes in the path from C_1 to C_3 which is the least common super concept of C1 and C2, and

N2 is the length given in the number of nodes on a path from C2 to C3.

N3 represents the global depth of the hierarchy and it serves as the scaling factor.

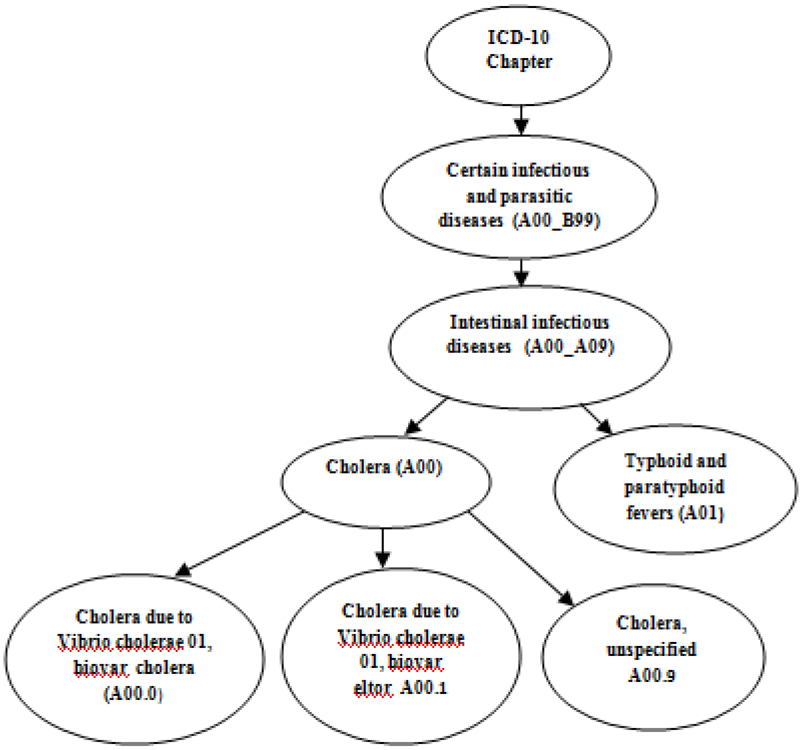

ICD-10 Chapter
Certain infectious and parasitic diseases (A00_B99)
Intestinal infectious diseases (A00_A09)
Cholera (A00)
Typhoid and paratyphoid fevers (A01)
Cholera due to Vibrio cholerae 01, biovar cholera (A00.0)
Cholera due to Vibrio cholerae 01, biovar eltor A00.1
Cholera, unspecified A00.9

Figure 1 fragment of Intestinal infectious diseases

For example from Figure 1: (LCS (A00.1, A00.9) = A00 and LCS(A00 ,A01) = A00_A09) of two concept nodes and N_1, N_2 are the path lengths from each concept node to LCS, respectively. 4.5 Al- Mubaid and Nguyen Similarity technique [5, 11] proposed measure take the depth of their least common subsume (LCS) and the distance of the shortest path between them. The higher similarity arises when the two concepts are in the lower level of the hierarchy. Their similarity measure is:

$$Sim (c_1, c_2) = \log_2 ([L(c_1, c_2) -1] \times [D- depth(L(c_1, c_2)] + 2) \qquad (2)$$

Where:

$L(c_1, c_2)$ is the shortest distance between c1 and c_2.

Depth $L(c_1, c_2)$ is depth of $L(c_1, c_2)$ using node counting.

$L(c_1, c_2)$ lowest common subsume of c_1 and c_2.

D is the maximum depth of the taxonomy.

The similarity equal 1, where two concepts nodes are in the same cluster/ontology. The maximum value of this measure occur when one of the concepts is the left most leaf node, and the other concept is the right leaf node in the tree. In the ICD-10 tree let us consider an example in ICD-10 terminology. The category tree is “Intestinal infectious diseases” and is assigned letter A in ICD10 terminology version 2016 at the link (http://apps.who.int/classifications/icd10/browse/2016/en#/A00-A09). This tree looks as follows:

Intestinal infectious diseases [A00-A09]
Cholera [A00]+
Typhoid and paratyphoid fevers [A01]+
Other salmonella infections [A02]+
Shigellosis [A03]+
Viral and other specified intestinal infections [A08]+
Other gastroenteritis and colitis of infectious and unspecified origin [A09]+

The similarity between “Cholera [A00]” and “Typhoid and paratyphoid fevers [A01]” is less similarity than the similarity between “Cholera due to Vibrio cholerae 01, biovar eltor [A00.1]” and “Cholera, unspecified [A00.9]”. However, in this measure they take into account the depth

The symbol "+" indicates that the concept can be further expanded into a sub tree (sub-concepts). For example, "Cholera" [A00] can be expanded to be as follows:

Cholera [A00]

Cholera due to Vibrio cholerae 01, biovar cholerae [A00.0]+

Cholera due to Vibrio cholerae 01, biovar eltor [A00.1]+

Cholera, unspecified [A00.9]+

of the LCS of two concepts, in the path length and leacock & chodorwo produce semantic similarity for two pairs [(A00, A01) and (A00.1, A00.9)] in sim (c_1, c_2) measure (Eq 2 in table 1) give high similarity in lower level in the ontology hierarchy ([A00.1, A00.3]).

Table 1: Measures Comparison

Pair of Concepts	P. L	L. C	C. K	Hisham Al-Mubaid & Nyguan Measure (Eq 2)
A00 – A01	0.37	2.13	0.91	3.2
A00.1 – A00.9	0.33	2.15	0.91	1.6

The higher numeric similarity result between (A00, A01) means the lower semantic similarity between them.

5. EVALUATION

5.1 Datasets:

There are no standard human rating sets for semantic similarity in biomedical domain. Thus, Hisham Al-Mubaid and Nguyen [3, 11] used dataset from Pedersen et. al [1], which was annotated by 3 physician and 9 medical index experts to evaluate their proposed measure in the biomedical domain.

Table 2 Dataset 1: 30 medical term pairs sorted in the order of the average [1].

Id	*Concept1*	*Concept2*	*Phys*	*Expert*	*Id*	*Concept1*	*Concept2*	*Phys*	*Expert*
4	Renal failure I12.0	Kidney failure I12.0	4.0000	4.0000	**27**	**Acne**	**Syringe**	**2.0000**	**1.0000**
5	Heart I51.5	Myocardium I51.5	3.3333	3.0000	12	Antibiotic (Z88.1)	Allergy (Z88.1)	1.6667	1.2222
1	Stroke I64	Infarct I64	3.0000	2.7778	**13**	**Cortisone**	**Total knee replacement**	**1.6667**	**1.0000**
7	Abortion O03	Miscarriage O03	3.0000	3.3333	**14**	**Pulmonary embolus**	**Myocardial infarction**	**1.6667**	**1.2222**
9	Delusion (F06.2)	Schizophrenia (F06.2)	3.0000	2.2222	16	Pulmonary Fibrosis (E84.0)	Lung Cancer (C34.1)	1.6667	1.4444
11	Congestive heart failure (I50.0)	Pulmonary edema (I50.1)	3.0000	1.4444	**6**	**Cholangiocarcinoma**	**Colonoscopy**	**1.3333**	**1.0000**
8	Metastasis (C77.0)	Adenocarcinoma (C08.9)	2.6667	1.7778	29	Lymphoid hyperplasia (K38.0)	Laryngeal Cancer (C32.0)	1.3333	1.0000
17	Calcification (M61)	Stenosis (H04.5)	2.6667	2.0000	21	Multiple Sclerosis (F06.8)	Psychosis (F06.8)	1.0000	1.0000
10	**Diarrhea**	**Stomach cramps**	**2.3333**	**1.3333**	22	Appendicitis (K35)	Osteoporosis (M80)	1.0000	1.0000
19	Mitral stenosis (I05.0)	Atrial fibrillation (I48)	2.3333	1.3333	23	Rectal polyp (K62.1)	Aorta (I70.0)	1.0000	1.0000
20	Chronic obstructive pulmonary disease (J44.9)	Lung infiltrates (J82)	2.0000	1.8889	24	Xerostomia (K11.7)	Alcoholic cirrhosis (K70.3)	1.0000	1.0000
2	Rheumatoid arthritis (M05.3)	Lupus (L93)	2.0000	1.1111	25	Peptic ulcer disease (K21.0)	Myopia (H52.1)	1.0000	1.0000
3	Brain tumor (G94.8)	Intracranial hemorrhage(I69.2)	2.0000	1.3333	26	Depression (F20.4)	Cellulitis (H60.1)	1.0000	1.0000
15	Carpal tunnel Syndrome (G56.0)	Osteoarthritis (M19.9)	2.0000	1.1111	**28**	**Varicose vein**	**Entire knee meniscus**	**1.0000**	**1.0000**
18	Diabetes mellitus (E10-E14)	Hypertension (I10-I15)	2.0000	1.0000	30	Hyperlipidemia (E78.0)	Metastasis (C77.0)	1.0000	1.0000

5.2 Experiments and Results

Table 2. Test set of 30 medical term pairs sorted in the order of the averaged physicians’ scores (taken from Pedersen et. al. 2005 [1]). Al-Mubaid and Nguyen [5, 11] find only 24 out of the 30 concept pairs in ICD-10 using http://apps.who.int/classifications/icd10/browse/2016/en browser version 2010.

Another biomedical dataset was used containing 36 MeSH term pairs [15]. The human scores in this dataset are the average evaluated scores of reliable doctors. UMLSKS browser was used [12] for SNOMED-CTterms, and MeSH Browser [13] for MeSH terms. Table 3, Table 4, Table 5, and Table 6 show Dataset2 along with human scores and scores of Path length, Wu and Palmer's, Leacock and Chodorow's, and Hisham Al-Mubaid & Nguyen techniques calculated using MeSH ontology. The term pairs in bold, in Table 3, Table 4, Table 5, and Table 6, are the ones that contain a term that was not found in MeSH Ontology and they were excluded from experiments.

Table3. Biomedical Dataset 2 (36 pairs) with human similarity scores (Human) and Path length's scores using MeSH ontology.

Id	Concept 1	Concept 2	Human	Path length
1	Anemia	Appendicitis	0.031	8
2	Meningitis	Tricuspid Atresia	0.031	8
.		.	.	.
.		.	.	.
36	Chicken Pox	Varicella	0.968	1

Table 4. Biomedical Dataset 2 (36 pairs) with human similarity scores (Human) and Wu and Palmer's scores using MeSH ontology.

Id	Concept 1	Concept 2	Human	Wu &Palmer
1	Anemia	Appendicitis	0.031	0.364
2	Meningitis	Tricuspid Atresia	0.031	0.364
.		.	.	.
.		.	.	.
36	Chicken Pox	Varicella	0.968	1.000

Table 5. Biomedical Dataset 2 (36 pairs) with human similarity scores (Human) and Leacock and Chodorow's scores using MeSH ontology.

Id	Concept 1	Concept 2	Human	Leacock & Chodorow
1	Anemia	Appendicitis	0.031	1.099
2	Meningitis	Tricuspid Atresia	0.031	1.099
.		.	.	.
36	Chicken Pox	Varicella	0.968	3.178

Table 6. Biomedical Dataset 2 (36 pairs) with human similarity scores (Human) and Hisham Al-Mubaid & Nguyen measure (SemDist) using MeSH ontology.

Id	Concept 1	Concept 2	Human	SemDist
1	Anemia	Appendicitis	0.031	4.263
2	Meningitis	Tricuspid Atresia	0.031	4.263
36	Chicken Pox	Varicella	0.968	0.000

6. CONCLUSIONS AND FUTURE WORK

In this paper we discussed the basics of semantic similarity techniques, the classification of single ontology similarity measures and cross ontologies similarity measures. We prepare a brief introduction of the various semantic similarity measures in biomedical domain. However, from all the above, we can used SemDist as semantic similarity measures in the biomedical domain. In future work, we intend to explore the semantic similarity techniques in the biomedical domain (ICD10, MeSH, and SNOMED-CT) within UMLS frame work. We also prepare implement a web-based user interface for all these semantic similarity techniques and to make it available freely to researchers over the Internet. That will be much helpful for interested researchers in the field of bioinformatics text mining.

7. REFERENCES

[1] Ted Pedersen, et al. " Measures of semantic similarity and relatedness in the biomedical domain ", Journal of Biomedical Informatics 40 (2007) 288–299.

[2] Hisham Al-Mubaid and Hoa A. Nguyen, “A Cluster-Based Approach for Semantic Similarity in the Biomedical Domain” Proceedings of the 28th IEEE, EMBS Annual International Conference New York City, USA, Aug 30-Sept 3, 2006.

[3] Hisham Al-mubaid & Hoa A. Nguyen “Measuring Semantic Similarity between Biomedical concepts within multiple ontologies” IEEE Trans Syst Man Cybern Part C: Appl Rev 2009, 39.

[4] Ahmad Kayed, et al. "Ontology Evaluation: Which Test to Use" 2013 5th International Conference on Computer Science and Information Technology (CSIT), IEEE, pp 45-48, 2013.

[5] Hisham Al-Mubaid and Hoa A. Nguyen, “New Ontology Based Semantic Similarity for the Biomedical Domain”, (2006) p 623 – 628.

[6] S. Anitha Elavarasi, et. al, “A Survey on Semantic Similarity Measure” International Journal of Research in Advent Technology, Vol.2, No.3, March 2014 E-ISSN: 2321-9637.

[7] Nguyen H., Al-Mubaid H. (2006) “New Semantic Similarity Techniques of Concepts applied in the biomedical domain and WordNet.” MS Thesis, University o f Houston Clear Lake, Houston, TX USA, 2006.

[8] World Health Organization, “International statistical classification of diseases and related health problems”. - 10th revision, edition 2010.

[9] Hisham Al-Mubaid and Hoa A. Nguyen, “Using MEDLINE as Standard Corpus for Measuring Semantic Similarity in the Biomedical Domain”, Sixth IEEE Symposium on BionInformatics and BioEngineering (BIBE'06), 2006.

[10] Mirjana Ivanovic& Zoran Budimac, An overview of ontologies and data resources in medical domains, Expert Systems with Applications 41 (2014) 5158–5166.

[11] Montserrat Batet Sanromà, “ontology-based semantic clustering”, PhD Thesis, 2010.

[12] UMLSKS. Available: http://umlsks.nlm.nih.gov

[13] MeSH Browser. Available: http://www.nlm.nih.gov/mesh/MBrowser.html