DOI: 10.7763/IPEDR. 2014. V72. 1

Productive Academic Database Enrolment

Sadi Evren SEKER[1], Mehmet Lutfi ARSLAN [1] and Menduh DINC [1] ·

[1] Istanbul Medeniyet University, Department of Business
{sadi.seker,lutfi.arslan,menduh.dinc}@medeniyet.edu.tr

Abstract. University libraries are, more than ever, confronting to allocate their resources in more efficient and productive ways. On the one hand, they try to diversify their options to have rich collections; on the other hand, they challenge the risk of purchasing expensive, repetitive and outdated service, which creates a dilemma. Universities and research institutes in terms of academic database selection are between the poor selection and the inefficient selection. In this study, we try to measure productivity of academic database selection of Turkish universities via statistical parameters, like the number of downloaded papers, the number of academicians, the number of students and so on. The results show that the efficiency varies and the research underlines the importance of productivity through a statistical parameterization of the competitive approach in the selection process. While doing this, we provide a decision support system for the online database selection for academic libraries.

Keywords: Strategic competition, Higher Learning, University Ranking, Online Databases, Research Indices, Decision Support System

1. Introduction

Online academic databases are one of the major parameters to indicate the research level of the universities. Beyond the number of selections, the efficiency and productivity of selection and the effective usage of online academic databases is also another important issue. According to study by Kaser in 2011, subscription to journals and databases represent about half of the budget of academic and special libraries. [1]

We aim to figure out a useful decision support system for the research institutes and universities during the enrollment. Also the decision support system is useful in bi-directional ways. In one direction the organizations can use the system during the decision phase of enrollment. On the contrary direction, the efficiency of enrollment can be figured out from the decision support system.

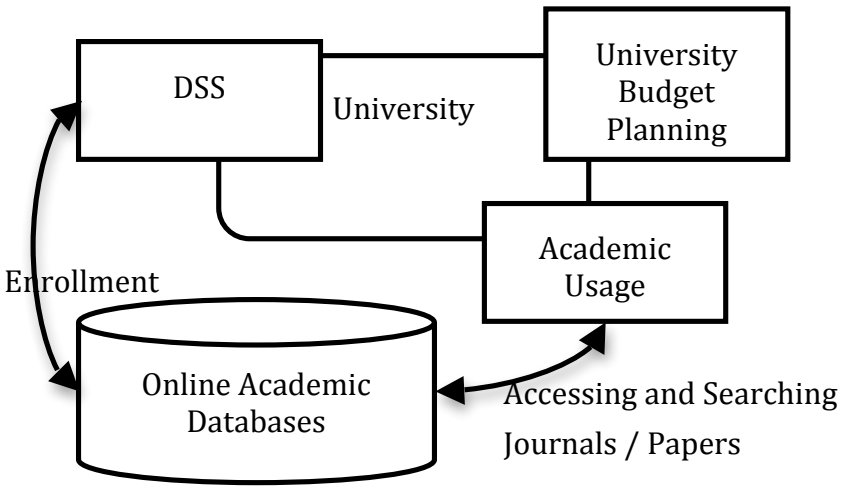


Fig. 1. Deployment of DSS

· Corresponding author. Tel.: + 9 0532 4467882
E-mail address: sadi.seker@medeniyet.edu.tr

As it is demonstrated on Fig. 1, universities have a decision phase during the enrollment of the online databases. The decision is mostly related with the university budget and the academic usage / demand to the online academic databases. If the university is enrolled in an online academic database, the university benefits from the database by searching or accessing journals / papers. Online academic databases have advantages to the search engines like well categorized and indexed paper access or stable information about the papers. Also the competition among universities has an encouraging effect over the decision.

During our research, we propose a new decision support system to the universities built on the efficiency of the enrollment. We have collected university effectiveness of online database enrollment for 6 years and we propose the decision support system over the online academic database prices, the number of papers accessed, library budget ratio, full time equivalent method on multiple databases like Proquest – ABI Inform Global, Passport, OECD Library, Emerald Management eJournals (EMeJ), Economist Intelligence Units, SAGE Premier Journals for Turkish universities. The collected data set is evaluated via artificial neural network (ANN). The output of decision support system is the prediction of the effectiveness of the online academic database enrollment

2. Background

The question of how university libraries will do their collecting in terms of opportunity cost is a relevant and contemporary one. While diversity of database services increases their options to have rich collections, on the other hand, they challenge the risk of purchasing expensive, repetitive and outdated services [1]. This challenge will surely last, as Dawid W. Lewis puts, "as large data sets become common with real-time ubiquitous data collection in many areas of science (often referred to as e-science) and the social sciences" [2].

Three ways to measure the value that academic libraries create are classified by Holt as "data by use (transactions), input (costs) and output (circulation, visitation)" [3]. In the nineties, SERVQUAL, a method measuring aspects of service quality and identifies, was used for assessment. Later, the focus of assessment shifted from service-oriented approach to user-oriented approach [4].

As university budgets shrank, productivity concerns emerged as the main factor in determining the value of academic libraries. ROI studies were to measure revenue-generating activities of libraries. In this regard, one example was the UIUC study by Roger Strouse in 2007, titled "ROI for Libraries Remains High." In this study, Strouse measured the value of a library in terms of revenue-generating activities [5]. Another study by Megan Oakleaf, The Value of Academic Libraries: A Comprehensive Research Review and Report, provided a review of methodologies and best practices in the field of ROI [6].

Yet, the increase in online content usage has become ROI studies less relevant. Since e-content is more quantifiable, libraries tend to develop holistic approaches to see the overall picture [7]. For example, Chicago's Saint Xavier University library managed its budget to eliminate inefficiencies, and an efficient mix of information resources [8].

As the variety of collections increase, the problem of allocating resources to provide productivity will become a dominant problem for university libraries. This study aims to view this problem from a cross-database and cost-oriented approach.

3. Online Academic Databases

This section covers the details of the data sets of online academic databases for Turkish universities.

Emerald Management eJournals (EMeJ) has 294 journals which 280 of them are peer reviewed journals, 1400 books and about 250 case studies.

Economic Intelligence Unit (EIU) publishes the statistical values about the universities for more than 200 products offered. The price of enrollment varies by the product and the number of member universities from Turkey is 6 by the end of 2012 and 5 by the end of 2013.

University	FTE	Budget	Usage	Cost
Anadolu Üniversitesi	20014	22,700.00 USD	3758	6.04 USD

Boğaziçi Üniversitesi	11294	22,700.00 USD	9754	2.33 USD
İstanbul Teknik Üniversitesi	24513	22,700.00 USD	6014	3.77 USD
İzmir Ekonomi Üniversitesi	6168	22,700.00 USD	3152	7.20 USD
Marmara Üniversitesi	46158	22,700.00 USD	6229	3.64 USD
Yeditepe Üniversitesi	17276	22,700.00 USD	0	0.00 USD

SourceOECD is an online academic database mainly supported by OECD (Organization for Economic Co-operation and Development) and this online academic database hosts 8890 books, 300 journals and 2700 conference proceedings. The 23 universities enrolled to SourceOECD is listed in the Table 2.

University	FTE	Budget	Usage	Cost
Adnan Menderes Üniversitesi	9304	11,810.81 USD	143	82.59 USD
Akdeniz Üniversitesi	12422	11,810.81 USD	350	33.75 USD
Aksaray Üniversitesi	1896	8,858.11 USD	53	167.13 USD
Anadolu Üniversitesi	20014	14,756.76 USD	910	16.22 USD
Bahçeşehir Üniversitesi	7882	8,858.11 USD	242	36.60 USD
Boğaziçi Üniversitesi	11294	11,810.81 USD	1104	10.70 USD
Çukurova Üniversitesi	24876	14,756.76 USD	329	44.85 USD
Dokuz Eylül Üniversitesi	36307	14,756.76 USD	907	16.27 USD
Ege Üniversitesi	30880	14,756.76 USD	386	38.23 USD
Fırat Üniversitesi	16284	11,810.81 USD	32	369.09 USD
Galatasaray Üniversitesi	2877	8,858.11 USD	212	41.78 USD
Gazi Üniversitesi	54791	14,756.76 USD	739	19.97 USD
Hacettepe Üniversitesi	30437	14,756.76 USD	801	18.42 USD
Hitit Üniversitesi	2776	7,381.76 USD	98	75.32 USD
İzmir Ekonomi Üniversitesi	6168	8,858.11 USD	175	50.62 USD
Karadeniz Teknik Üniversitesi	30512	14,756.76 USD	243	60.73 USD
Koç Üniversitesi	4124	8,858.11 USD	249	35.57 USD
Kocaeli Üniversitesi	27501	14,756.76 USD	117	126.13 USD
Marmara Üniversitesi	46158	14,756.76 USD	582	25.36 USD
Mersin Üniversitesi	11487	11,810.81 USD	115	102.70 USD
Muğla Sıtkı Koçman Üniversitesi	11976	11,810.81 USD	304	38.85 USD
Orta Doğu Teknik Üniversitesi	25313	14,756.76 USD	1061	13.91 USD
Uludağ Üniversitesi	29683	14,756.76 USD	124	119.01 USD

Passport GMID is an online data source holding information about sectors, countries and customers. There are more than 16.000 market research from 80 different countries and more than 8 million statistical value customer behavior and predictions until 2020 for 83 country and 7 regions. The university enrollment from Turkey to the Passport is given in Table 3.

University	FTE	Budget	Use	Cost
Atılım Üniversitesi	5292	20,270.27 USD	159	127.49 USD
Boğaziçi Üniversitesi	11294	20,878.38 USD	3865	5.40 USD
İhsan Doğramacı Bilkent Üniversitesi	10673	20,270.27 USD	1875	10.81 USD

İstanbul Teknik Üniversitesi	24513	20,270.27 USD	5921	3.42 USD
Koç Üniversitesi	4124	20,270.27 USD	4346	4.66 USD
Orta Doğu Teknik Üniversitesi	25313	20,270.27 USD	1203	16.85 USD
Özyeğin Üniversitesi	953	20,270.27 USD	8560	2.37 USD
Pamukkale Üniversitesi	18767	4,054.05 USD	25	162.16 USD
Sabancı Üniversitesi Bilgi Merkezi	3812	20,270.27 USD	7758	2.61 USD

Other databases we have researched are EBSCOHost, JSTOR Business 1,2,3, Gale Expanded Academic ASAP and SAGE Premier Journals, Proquest – ABI Inform Global. We can not give the data sources and full tables because of the size limitations of the paper but we use the full values in the construction of decision support system.

4. Decision Support System

In order to construct a decision support system, the data set collected is processed via artificial neural network (ANN). During the deployment of the decision support system, the feature extraction has an important role. The features plays a base model role for the decision support is extracting using some statistical methods. In this section, the feature extraction methods are explained just before the details of the DSS methods as a business intelligence method.

Full Time-Equivalent (FTE) method: Method is mainly designed to compare values from different domains. For example the contribution of two employee in different departments or the comparison of the contribution of publishing articles or lecturing are values from different domains. In order to compare those values from different domains, the FTE models a common domain where both values can be converted. During the conversion, each value has a weight coefficient depending on their importance. For example in a university, if publishing articles is more important than lecturing courses, than the coefficient of article publish would be 1 while the coefficient of lecturing is 0.5. In this model the number of articles published would be two times weighted than the number of courses lectured.

The similar approach is implemented by the online academic databases. The weight of universities with higher number of students or academicians, pay more for the enrollment. We use the FTE value as a feature but our research also underlines that the FTE parameter is misdirection for the decision. In our decision support system we propose a decision mainly built on the number of paper usage rather than the FTE.

Artificial Neural Network (ANN) The decision support system is built by using artificial neural network. The purpose of ANN studies is adapting the biological neural networks into data processing. Multi-layer perceptron (MLP) is a developed version of ANN, which organizes the neurons into layers as input, output or hidden layers. Fig. 2 demonstrates the approach in a data flow between layers[10]. The ANN model has been one of the attractive tools used in geo-engineering applications due to its high performance in the modeling of non-linear multi-variate problems. Hecht-Nielsen [11] and Schalkoff [12] indicate that an ANN may be defined as a structure comprised of densely interconnected adaptive simple processing elements that are capable of performing massively parallel computations for data processing and knowledge representation.

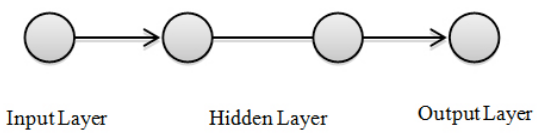


Fig. 2. Data flow between layers

The input layer is directly connected to the inputs which will be processed within the system. The hidden layer can hold more than one layer depending on the complexity of the system and finally the output layer holds the results. Fig. 2 can be detailed if each of the layers is demonstrated by a real number of neurons. Fig. 3 is a detailed version of Fig. 2 with multiple neurons in each of the layers. Please note that each neuron on a layer is directly connected to all the neurons on the next layer.

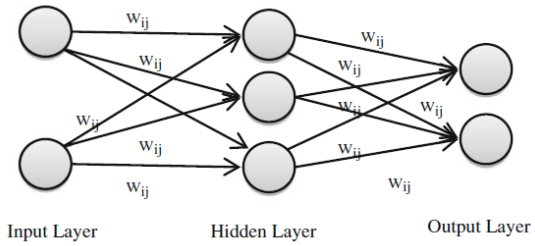


Fig. 3. Generic view of MLP with multiple input and weight values of the synapses

Each of the connections in Fig. 3, which are called a synapsis in a biological view, has a weight value, which is a coefficient to the value on the neuron which is indicated by "wij", where i and j indices are the weight value between the neurons i and j [13]

The aim of this study is to calculate the best possible values of the above weights. During this calculation the back propagation algorithm is implemented. The algorithm changes the weight values on the synapses by the result achieved on the output layer.

Most of the ANN systems use the sigmoid function for the decision of firing the neuron [14]. The generic formula of the sigmoid function is given in equation (1). Mitchell uses the word "logistic function" and the "sigmoid function" synonymously. He also calls this function a "squashing function" and the sigmoid (a.k.a. logistic) function is used to compress the outputs of the "neurons" in multi-layer neural nets [15]. The generic two dimensional view of the function is given in Fig. 4.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

Besides the above ANN information, during this study the two parameters of ANN have crucial roles which are learning rate and the momentum factor.

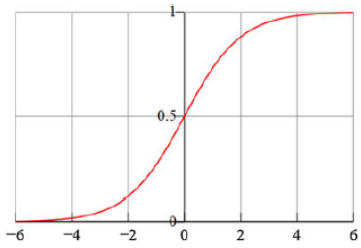


Fig. 4. Visualization of sigmoid function

The learning rate of the ANN is denoted by α, which varies from 0 to 1. The learning rate is a parameter to control the amount of weight adjustment at each step of training to determine the learning rate at each step. It affects the convergence of back propagation network (BPN). A large value of α indicates the faster learning with high probability of error, while the lower learning rate indicates slower but robust learning.

A second crucial parameter of ANN is the momentum, which is denoted by η and varies from 0 to 1. The momentum is used to speed up the process. BPN has two major disadvantages which are larger training time and slow convergence. The momentum parameter helps to improve the convergence and the training time while reducing the oscillation of learning.

The ANN shows a great success on the predictive model for the decision support system. The model is also useful for the university based decision support on the digital library enrollment as well as the university rankings [16]. The correlation among university reputation and the productive library enrollment can also be a future work based on the correlation methods [17]. We also plan to extend our research based on the web based parameters and user comments about the academic libraries[18],[19] based on Turkish texts [20].

5. Conclusion

This study first time underlines the efficiency of online academic database enrollment. The research is built on 10 different database statistics for the universities from Turkey. The statistics are collected based on the budget, FTE, cost and number of downloads for each of the database and each of the university. We have proposed a decision support system, which is built on artificial neural network.

6. References

[1] D. Kaser, “On Average: How Your Library Budget Stacks Up”, Computers in Libraries, Vol. 31 No. 2, pp. 33-5, 2011

[2] D. Hazen, “Rethinking research library collections”, Library Resources & Technical Services, 54(2), pp. 115-121, 2010.

[3] D. W. Lewis, “A strategy for academic libraries in the first quarter of the 21st century”, College & Research Libraries, 68(5), pp. 418-434, 2007.

[4] L. E. Holt, Performance management in Library Success : a celebration of library innovation, adaptation and problem solving, Ebsco Publishing, p. 128, 2006.

[5] Xi Shi and Sarah Levy, “A Theory-Guided Approach to Library Services Assessment,” College & Research Libraries 66 no. 3, pp. 266-277, 2005. Available online at http://crl.acrl.org/content /66/3/266.full.pdf+html [Accessed 13 February 2014].

[6] Roger Strouse, “ROI for Libraries Remains High,” Access: Asia’s Newspaper on Electronic Information Products and Services, 63, 2007. Available online at: http://www.aardvarknet. info/access/number63/monthnews.cfm?monthnews=06 [Accessed 13 February 2014].

[7] Megan Oakleaf, “The Value of Academic Libraries: A Comprehensive Research Review and Report”, Association of College and Research Libraries: Chicago, 2010.

[8] P.Kaufman & S.B. Watstein, “ Library value (return on investment, roi)and the challenge of placing a value on public services”, Reference Services Review, 36(3), pp. 226□331, 2008.

[9] J. P. Kusik & M. A. Vargas, Implementing a "holistic" approach to collection development, Library Leadership & Management, 23(4), pp. 186□192, 2009.

[10] I. Ocak, S. E. SEKER (2012), Estimation of Elastic Modulus of Intact Rocks by Artificial Neural Network, Rock Mechanics and Rock Engineering, Springer, Vol. 45, Is. 6, pp. 1047 – 1054, DOI: 10.1007/s00603-012-0236-z, Nov. 2012

[11] Hecht-Nielsen R (1987) Kolmogorov’s mapping neural network existence theorem. Proceedings of the first international conference on neural networks: 11-14 San Diego, USA.

[12] Schalkoff RJ (1997) Artificial neural network. New York, USA.

[13] Wasserman PD, Schwartz T (1988) Neural networks II. What are they and why is everybody so interested in them now?. IEEE Expert 3(1):10-15.

[14] Mitchell TM (1997) Machine Learning, New York, USA.

[15] S. E. SEKER, Y. Unal, Z. Erdem, H. Erdinc Kocer (2014), “Ensembled Correlation between Liver Analysis Outputs”, International Journal of Biology and Biomedical Engineering, ISSN: 1998-4510, Volume 8, pp. 1-5, 2014

[16] Mehmet Lutfi ARSLAN, Sadi Evren SEKER (2014), ”Web Based Reputation Index of Turkish Universities”, International Journal of E-Education E-Business E-Management and E-Learning (IJEEEE), 2014, Issn : 2010-3654, vol.4, is.3, pp.197-203

[17] Sadi Evren SEKER, Cihan MERT, Khaled Al-Naami, Ugur AYAN, Nuri OZALP, “Ensemble classification over stock market time series and economy news“, Intelligence and Security Informatics (ISI), Proceeding of 2013 IEEE International Conference, pp 272 – 273, ISBN 978-1-4673-6214-6

[18] Sadi Evren SEKER, Khaled Al-NAAMI, Latifur KHAN, “ Author Attribution on Streaming Data“,Information Reuse and Integration (IRI), 2013 IEEE 14th International Conference on , IEEE IRI pp. 497 – 503, Aug. 2013

[19] Sadi Evren SEKER, Khaled Al-NAAMI “Sentimental Analysis on Turkish Blogs via Ensemble Classifier“, PROCEEDINGS OF THE 2013 INTERNATIONAL CONFERENCE ON DATA MINING, ISBN:1-60132-239-9, DMIN, pp. 10-16, 2013

[20] Sadi Evren SEKER, Banu DIRI (2010) , “TimeML and Turkish Temporal Logic” ICAI 2010, WolrdComp 2010,(July 12-15, 2010, USA)