Automated Ranking of E-commerce Product Offers Based on a TQM Approach

A Case Study in the Computer Networking E-commerce Products

Ion POPA
The Bucharest Academy of Economic Studies, Romania
E-mail: popaion_2000@ yahoo.com
Camelia Elena CIOLAC
The Bucharest Academy of Economic Studies, Romania
E-mail: ciolac_c@ yahoo.co.uk
Sebastian CEPTUREANU
The Bucharest Academy of Economic Studies, Romania
E-mail: ceptureanu@ yahoo.com
Eduard CEPTUREANU
The Bucharest Academy of Economic Studies, Romania
E-mail: eduard_ceptureanu@ yahoo.com

Abstract

In the knowledge-based economy , it is the knowledge about market offer and products' quality that assures the e-Procurement process's performance. A simple price comparison or a juxtaposition of technical features for a small set of products is insufficient for a good investment decision , especially regarding the IT hardware field. The paper aims to propose an automated web content mining framework to classify e-commerce products based on a Total Quality Management (TQM) approach. The main beneficiary of this framework include: companies that must adopt an IT investment decision considering the quality-price trade-off, IT providers that could identify opportunities for strategic alliances based on product quality ressemblance , IT consultant etc.

Keywords: *e-commerce, software agent, knowledge discovery, TQM STEM method, IT investments*

JEL classification: D8, L2, M1

1. Introduction

The paper aims to propose an automated web content mining framework to classify e-commerce products based on a Total Quality Management (TQM) approach.

The knowledge discovered using speciallized software agents is processed with TQM algorithms and the result can constitute a valuable instrument in the e-Procurement decisions .All resulting knowledge products are published in a web knowledge portal and can be accessed by all stakeholders:

- customers that want to found their investment decisions on rigurous reasoning
- companies that want to take an overview of the market's dynamics and of the competitors' capabilities and products in order to found the company's medium and long-term strategy

This innovative perspective upon e-commerce brings many advantages that include but are not restricted to the following:

- facilitates horizontal integration among e-commerce websites that advertise non-disjunctive sets of products
- provides a quality assesment tool to evaluate e-commerce offers
- represents an innovative quantitative model to support the decision-making process in SCM (supply-chain management)

In order to prove the benefits of this innovative approach , we carried out a case study in on the Romania IT electronic market , precisely an automatic ranking of network switches based on the TQM approach.

2. Preliminary considerations

Nowadays most of e-commerce we bites provide customers with tools for products' comparison. Still, this approach has some disadvantages as it implies that the customer would have a-priori knowledge about the importance of technical features as well as an exhaustive knowledge of e-commerce products to be able to include them in the comparison process .

It is in these areas that the framework we propose shows its strenght, as it automizes the web content mining process of e-commerce websites by use of speciallized software agents .

The software information agents are trained to discover knowledge from a variety of e-commerce websites and this approach results in an enhancement of the set of products that undergo the evaluation process.

The agents store the discovered products in an unstructured manner at first, because it is important to register all the technical features that characterize a product type. Therefore, the first stage of the framework consists in registering all information available on different websites for the analyzed product type.

The target websites that are subject to web content mining are the e-commerce websites that advertise ITC hardware, including telecommunications and networking products.

Further analysis is carried out on the stored data, by the use of set and matrix computation. Therefore the second part of the framework we propose consists of a quantitative model for e-commerce products ranking.

3. Framework for web content mining

The general framework for automated ranking of e-commerce product offers based on a TQM approach is presented in figure 1 and consists of two main steps:

S1) Knowledge discovery

S2) Products' ranking using TQM techniques

Let us consider the analysis for products of generic type T in the framework description, where T can take values in the set „network switch", „router", „video board", „laptop" etc.

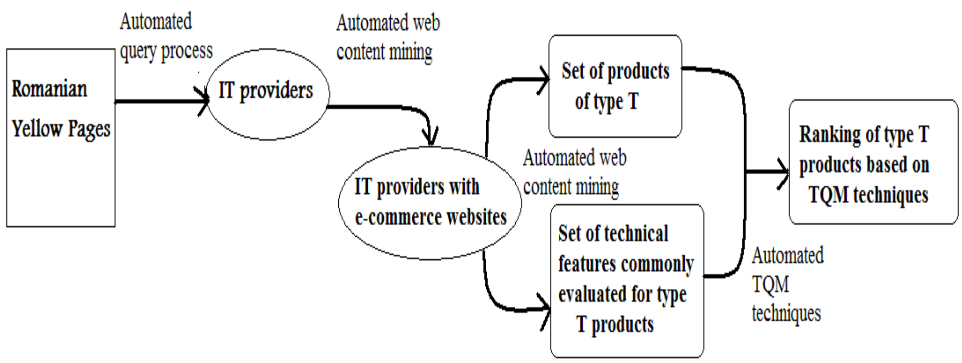


Figure 1 The general framework for automated ranking of e-commerce product offers

S1) Knowledge discovery through web content mining

One of the good practices in the e-Procurement process refer to the selection of potential providers using important criteria such as: historic contacts, market share, brand reputation etc.

Using the web version of Yellow Pages, a comprehensive list of all IT providers activating in a certain geographical area can be retrieved.

Let

P={ pi | pi is IT provider listed in the Romanian Yellow Pages }

The list contains, along other contact information, links that point to the IT proviers' websites.

A speciallized agent can iterate through the list and connect to each IT provider *pi*'s (*i=1..|P|*) website in order to determine whether the provider offers e-commerce services as well. This task can be acomplished by parsing the provider's website content in search of keywords like „basket", „login", "add to basket" ecc.

Let

Pec= { pi | pi P and pi offers e-commerce services}

All IT providers’ websites that offer e-commerce services are stored in the „T_Providers” table of the database as tuples of (URL, Provider_Name, Provider_Reputation). The provider *pi*’s reputation is assesed as a score on a 1 to 5 scale, using the historic experience bound to that provider as well as other customers’ experience obtained by **Opinion Mining** on the specialized IT on-line forums.

E-commerce websites use a variety of templates for presenting their product offers. Given the non-uniformity of products classification among e-commerce websites, the speciallized software agents are trained to identify the search input form within the website. Therefore, the software agent locates the <input type=”text” > and the <input type=”submit”> HTML tags within the source code of the e-commerce home page. The automate search form filling and submission results in a set of catalog entries for the specific produc type T.

Let

Xi={ xij | xij is product of type T retrieved from provider pi’s e-commerce website }

For each element xij in the Xi set, the software agent opens the HTML page corresponding to xij’s product details page. Two situations are met: in the first one, product technical specifications are listed directly in the product details page, while in some cases, the product details page provides a link to a separate file of technical specifications. Either way, the software agent searches for the „technical specifications” section and retrives Sij, where:

Sij={ (fk , vijk) | fk is a technical feature , vijk is the value of metric fk for product xij with xij □ Xi }

At the end of the web content mining process, some further processing is carried out on the discovered knowledge.

Let **$X = U_{i=1,|Pec|}$ Xi** be the total set of products that undergo the analysis. We consider the projection function **$pr_f s$= fk where s=(fk,vijk) □ Sij** the function that retrieves the technical feature coordinate from a tuple (technical feature, hardware product).

Let **$S= \cap_{i=1,|P|\,;\,j=1,|Xi|}\ pr_f\ Sij$**

Be the intersection of all features of different Sij. Consequently, S contains the set of technical features that are measured for all products of type T advertised through Romanian e-commerce websites. It is important to consider only the commonly assesed technical features for the products that undergo our analysis because the lack of values is not allowed in the TQM techniques we are going to use.

S2) Products' ranking using TQM techniques

In order to rank the products of type T, we shall apply the STEM method described in Ref [1].

The STEM method is a TQM technique that uses the importance of different parameters in order to compare a set of products in terms of quality.

- The matrix M is built as a 2-D collection of dimensional data .In our analysis, the rows represent the technical parameters of set S and the columns store the products' values for these.

Let us consider the projection function **pr** $_{fk}$ **s= vijk , where s=(fk,vijk)** □ **Sij**

Therefore , we can formalize **M[i1][j1]= pr** $_{fi1}$ **sj1**

- The second stage of the STEM method consists in elaborating an „importance matrix" to asses the relative importance between all possible pairs of technical features in set S , that is **(si , sj)** □ **S x S** with i,j=1..|S|. For each product type T an „importance matrix" *I* is built. In order to obtain reliable results, we suggest using a Delphi research among IT specialists and practitioners to accurately asses the importance of different technical parameters of type T products into practice.

- Based on the matrixes M and I we are able to determine the absolute technical level Ref [1]:

$$\mathbf{ATLj = 1000 * \prod (M[i][j] / M[i][jref])^{h(i)} \quad with \ \ i=1,|S| \ , j=1,|Pec|}$$

where h(i) represents the relative importance of feature in the total set of technical features S and jref represents a product of type T that belongs to $U_{i=1,|Pec|}$ Xi and that is considered as a reference or comparison base. In some cases, in order to be equidistant to all products that undergo the analysis , the jref product that is considered the reference for comparison can be replaced with the product of type T that is currently in use and needs to be upgraded / replaced .

- Based on all computed ALTj j=1,|Pec| , the STEM algorithm allows us to rank products according to the following formula as described in Ref [1]:

$$\mathbf{RTLj= (ALTj / (max\ ALTj)) * 100}$$

The challenges in running the STEM algorithm on IT hardware products resides in the literal description of some technical features, that cannot be subject to numeric computation.

Therefore, this paper proposes an extension of the TQM STEM methodology, considering the following aspects:

- hardware products are characterized by standards
- hardware products do not operate separately and therefore an important factor referees to the interfaces and connector options.

The difficulty that arises refers to the possibility of quantifying these features.

In order to asses the compatibility, adaptability and interoperability of different hardware products we propose the following model:

Let Nf = the number of interfaces of the analyzed product

For example, for a switch 24x RJ-45 2xMini-GBIC we have Nf=26. The variety in the interface types is of the same importance, because even if used for configuration, communication or device plug&play, they provide ways of interacting and connecting with other hardware devices.

Next, we consider the number of standards Ns implemented in the device. This metric is very important because the higher Ns is, the higher are the chances of the hardware product interoperating with other IT products.

Another important metric that should be considered is the number Np of protocols supported by the hardware device. Consequently,

Np = Nop + Nmp

where Nop is the number of operational protocols and Nmp represents the number of management protocols supported by the analyzed hardware.

Using the above described formalization, we can quantify the descriptive qualitative technical features of hardware products with respect with TQM STEM's general guidelines.

4. A Case study in Networking Switching Products

In order to prove the applicability of the proposed framework, let us consider the networking switching products. The web content mining step of the proposed framework high lightens the set S of the commonly evaluated technical features for network switch products :

- Number of ports => Nf
- Minimum Ethernet Data Transfer Rate
- Maximum Ethernet Data Transfer Rate
- Data RAM Buffer size
- MAC Address Table Space
- Power consumption
- Physical dimensions =>Volume in space
- List of Standards implemented => Ns
- List of Protocols supported => Nop , Nmp

Given this set of technical features we build the matrix M according to STEM methodology, using an explanatory dataset, presented in table 1.

An example dataset of technical features for different switch products

Table 1

Feature	**Dim**	**Sw1**	**Sw2**	**Sw3**	**Sw4**	**Sw5**	**Sw6**	**Sw7**
Nf	-	26	26	24	24	16	8	5
Min Eth Data Transf Rate	Mbps	10	10	10	10	10	10	10
Max Eth Data Transf Rate	Mbps	1000	1000	100	100	1000	1000	100
Data RAM Buffer size	KB	512	256	512	256	128	128	64
MAC Address Table Space	KB	32000	8000	8000	8000	8000	4000	1000
Power consumption	W	100	100	10	10	40	10.5	2.5
Volume in space	mm^3	6784.5	2976.3	6219.1	2710.4	4065.6	866.4	243.7
Ns	-	4	4	7	3	3	4	3
Nop	-	1	1	1	1	1	1	1
Nm	-	0	0	0	0	0	0	0

Switches Swi having i=,7 are hardware devices advertised through Romanian IT e-commerce websites , but we preferred to code them because the analysis focuses at this point on technical features and not on brand names. Also, we considered the analysis for a set of unmanaged switches (therefore Nmp = 0 and will not be included in the computation)

The "importance matrix" I identified by networking specialists is presented in table 2.

In table 2, cell values have the following meaning
, according to STEM methodology presented in Ref [1]:

$$I[j][k] = \begin{cases} 0 \text{ if feature j is less important than feature k} \\ 1 \text{ if features j and k have simillar importance} \\ 2 \text{ if feature j is more important than feature k} \\ 4 \text{ if feature j is far more important than feature k} \end{cases}$$

The absolute technical levels for the analyzed switches, having Sw7 as a comparison reference , for example, are :

$$ATL\ 1 = 5.2^{0.144} + 1^{0.02} + 10^{0.09} + 8^{0.18} + 32^{0.162} + 40^{0.018} + 27.83^{0.009} + 1.33^{0.189} + 1^{0.189} = \mathit{10.837}$$
$$ATL\ 2 = 5.2^{0.144} + 1^{0.02} + 10^{0.09} + 4^{0.18} + 8^{0.162} + 40^{0.018} + 12.21^{0.009} + 1.33^{0.189} + 1^{0.189} = \mathit{10.307}$$
$$ATL\ 3 = 4.8^{0.144} + 1^{0.02} + 1^{0.09} + 8^{0.18} + 8^{0.162} + 4^{0.018} + 25.51^{0.009} + 2.33^{0.189} + 1^{0.189} = \mathit{10.319}$$
$$ATL\ 4 = 4.8^{0.144} + 1^{0.02} + 1^{0.09} + 4^{0.18} + 8^{0.162} + 4^{0.018} + 11.12^{0.009} + 1^{0.189} + 1^{0.189} = \mathit{9.97}$$
$$ATL\ 5 = 3.2^{0.144} + 1^{0.02} + 10^{0.09} + 2^{0.18} + 8^{0.162} + 16^{0.018} + 16.68^{0.009} + 1^{0.189} + 1^{0.189} = \mathit{10.017}$$
$$ATL\ 6 = 1.6^{0.144} + 1^{0.02} + 10^{0.09} + 2^{0.18} + 4^{0.162} + 4.2^{0.018} + 3.55^{0.009} + 1.33^{0.189} + 1^{0.189} = \mathit{9.767}$$

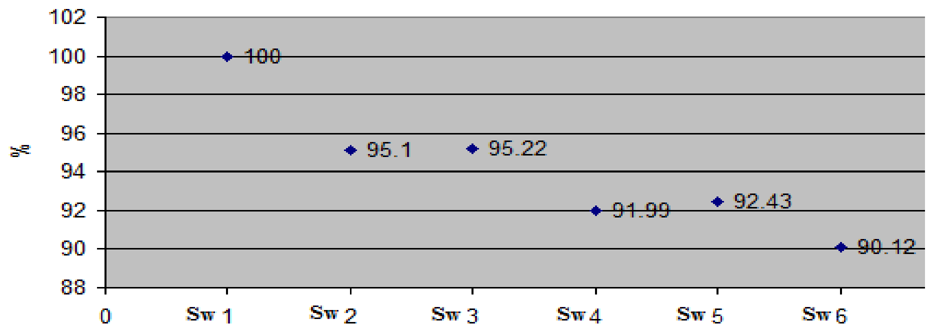
Importance matrix for network switch technical features

Table 2

	Nf	**Min Eth Data Transf Rate**	**Max Eth Data Transf Rate**	**Data RAM Buffer size**	**MAC Address Table Space**	**Power consumption**	**Volume in space**	**Ns**	**Nop**	**Line sum**	**h(i)**
Nf	-	4	2	1	1	4	4	0	0	*16*	*0.144*
Min Eth Data Transf Rate	0	-	0	0	0	1	2	0	0	*3*	*0.02*
Max Eth Data Transf Rate	0	2	-	0	1	4	4	0	0	*11*	*0.09*
Data RAM Buffer size	1	4	4	-	1	4	4	1	1	*20*	*0.18*
MAC Address Table Space	1	4	1	1	-	4	4	1	1	*18*	*0.162*
Power consumption	0	1	0	0	0	-	1	0	0	*1*	*0.018*
Volume in space	0	0	0	0	0	1	-	0	0	*0*	*0.009*
Ns	2	4	4	1	1	4	4	-	1	*21*	*0.189*
Nop	4	4	2	1	1	4	4	1	-	*21*	*0.189*

Therefore, compared to Sw7, we obtain the following relative technical levels for the analyzed products:

RTL 1 = 100%, RTL 2 = 95.10%, RTL 3 = 95.22%, RTL 4 =91.99%, RTL 5 = 92.43%, RTL 6 = 90.12%



Nf :	26	26	24	24	16	8
Max Eth Speed(Mbps)	1000	1000	100	100	1000	1000
Data RAM Buffer(KB)	512	256	512	256	128	128
MAC Table Spac(KB)	32000	8000	8000	8000	8000	4000
Ns	4	4	7	3	3	4

Network Switch

Figure 2 Rank for e-commerce products of type „network switch" and the differenciating parameters

According to TQM theory, quality classes group products whose relative technical levels vary with less than 3%-4% from each other.

In this case, the network switches have been groupped in 3 quality classes:

Class 1: Contains only Sw1 and is characterized by outstanding MAC table space and Data buffers as well as a large number of ports.

Class 2: Contains switches Sw2, Sw3. This class is characterized by a large number of ports and a trad-off between the maximum speed versus the Data RAM buffer size. A large number of communication protocols supported as well as the implementation of standards include a product in this quality class even if its operative performance metrics are of medium values.

Class 3: Contains switches Sw4, Sw5, Sw6. This class enlightens the trade-off between the number of ports and the maximum speed provided on each port and is characterized by smaller Data RAM buffer size.

In conclusion, the automatic e-commerce product ranking framework that we propose can become a useful instrument to assist the IT investing decision, based on a Total Quality Management approach.

References

[1] Ionescu S.C. – *Quality Management*, Bucharest: Didactic and Pedagogic Press, 2000

[2] Nicolescu O., Verboncu I. – *Management*, Bucharest: Universitara Press, 2008

[3] Nonaka, H. - *The Knowledge Creating Company*, Oxford: Oxford University Press, 1995

[4] Popa, Ion – *Strategic management*, Bucharest: Economic Press, 2004

[5] Raynolds Janice, Mofazali Roya - *The Complete E-Commerce Book: Design, Build & Maintain a Successful Web-based Business*, Copyrighted material, www.amazon.com

[6] Reynolds, Janice - *Complete e-commerce book: design, build [and] maintain a successful, The web-based business*, CMP Books Press, San Francisco, 2004

[7] Loton, T. - *Web Content Mining With Java,* England: John Wiley & Sons, 2002

[8] http://www.yellowpages.ro/cauta/bucuresti/computer.html

[9] www.dc-shop.ro

[10] www.emag.ro