IMPROVING THE SEARCH AND CATALOGUING OF ITEMS IN C2C E-COMMERCE PORTALS

Antonio Gallardo, Jose Jesus Castro-Schez, Milagros Hazas
Escuela Superior de Informática, Universidad de Castilla-La Mancha, 13071, Ciudad Real, Spain

Juan Moreno-Garcia
Escuela Universitaria de Ingenieía Técnica Industrial, Universidad de Castilla-La Mancha, 45004, Toledo, Spain

Keywords: Fuzzy cataloguing, fuzzy search and e-commerce.

Abstract: The business achievement among consumer via e-commerce is getting more important at the present time. In this paper, we propose to make use of fuzzy logic with the aim to improve the search and cataloguing of goods and services in Consumer-to-Consumer electronic commerce (E-commerce) portals (e.g. ebay). These portals are the media through most the electronic transactions among consumers are conduced today. We suggest a method that tries to adapt to users' real needs. It allows to buyer carry out searches in an imprecise way and to the seller to deal with catalogues of items (goods or services) described also in a lacking exactness way.

1 INTRODUCTION

In the last decade, the coming into sight and consolidation of the World Wide Web, that involves the establishment of a competitive environment, has forced to the firms to develop their sites to a high level of sophistication and integration (Kowtha and Choon, 2001; Laudon and Laudon, 2005) even allowing the achievement of business all around the world (Turban et al., 2000). This fact has caused the appearance of a new infrastructure to this new business paradigm known as *electronic commerce* (e-commerce). It can be defined as any business that is transacted electronically (Cameron, 1997). The e-commerce processes and technologies have introduced new ways of doing business (Chaudhury and Kuilboer, 2001).

Consumer-to-consumer (C2C) e-commerce relates to any business where the transaction occurs between two consumer who negotiate trying to reach an agreement or compromise. In spite of the majority of the commercial transactions are made in traditional ways yet, many consumers and firms which use Internet to perform commercial activities are appearing and they are obtaining benefits they could not obtain using traditional ways. Nowadays, C2C e-commerce portals are reaching a big summit but the full development of these portals and their use needs the development and/or integration of methods which facilitate the search and inclusion of items (goods or services) in catalogues. In this way, the accomplishment of commercial activities between consumer will be strengthen.

The great number of portals dedicated to C2C e-commerce use lexicographic objects arrangements and searches (i.e. mp3 player, laptop) or direct searches of the desired commercial model (i.e. iPod nano, Vaio VGN-SZ2HP/B). Once, a set of items have been found in the catalogue, they usually allow to give a order to this set by means of the specification of the values of some characteristics (e.g. price).

These portals do not allow to us the search giving only a description of the item by means of the specification of a set of characteristics, theirs values, the importance or relevance of these characteristics in our search and even less do it in an imprecise way.

The lexical search has the problem that if the seller does not mention in the item textual description an important aspect when he includes it into the catalogue, it will be more complex, difficult and possibly unsuccessfully the search. Other characteristic which defines these kinds of searches and arrangements is the use of synonymous of the entry given by the user. Then the search engine could find and show items whose category is totally different with the category of the item that user is searching. This problem appears because in the textual description of these items

Gallardo A., Jesus Castro-Schez J., Hazas M. and Moreno-Garcia J. (2007).
IMPROVING THE SEARCH AND CATALOGUING OF ITEMS IN C2C E-COMMERCE PORTALS.
In *Proceedings of the Ninth International Conference on Enterprise Information Systems - SAIC*, pages 13-19
DOI: 10.5220/0002353300130019

have been used synonymous. This could cause that user do not find quickly what he really wants and he leaves the search. In this kind of portal stand up *ebay* (Ebay,), which is the most used, or *froogle* (Froogle,), which has a big search engine.

Others portals, as the Spanish *compras* (Compras,), add improvements such as the items arrangement by mean of a common and important characteristic and the search by mean of a filter which is usually the price.

The search based on commercial models force to the user to have a high knowledge of the market of the item category. He must enter the model that he desires. This is the most precise method of search, but only a little group of users could use this portal successfully. This fact decrease the number of possible buyers and makes the portal less attractive, so the amount of sales the portal could obtain will decrease.

In this paper, we propose to make use of the fuzzy logic to allow to the user describe items giving theirs characteristics and values in an imprecise way and we suggest a method to make searches more precise. It has been designed to satisfy the users' needs with the aim to carry out searches more extensive and detailed.

The remainder of the paper is organized as follow. Section 2 shows how is the interaction with the portal and the users' requirements with regard to the search and cataloguing of items. In Section 3 how to catalogue an item and how the search process is carried out will be detailed. An example will be presented in Section 4. Finally, the conclusions obtained during the design and development of this portal will be presented.

2 PORTAL REQUIREMENTS

The portal suggested in this paper will be used in C2C e-commerce. It will allow to catalogue and search items through the specification of their characteristics and theirs values given in a imprecise way. Also, we allow to associate information about the relevance of each characteristic according with user's preferences. In this way, we try to obtain a list of items arranged that will be given to the user as a recommendation. This allow to the buyers to find easier what they are searching.

To establish the users' requirements, first we should determine what users are going to interact with the portal and how they will do. In this sense it can be distinguish three different users:

- *Administrator:* He manages and defines the different items categories and the characteristics of each category. Administrator should check that all categories are well determined and that the characteristics which define to each category will be useful to distinguish the items belonging to that category. An important part of the portal success depends on it.
- *Seller:* He enters items into the catalogue which could be obtain in exchange for payment by buyers. Seller will describe each item by means of a characteristics set which will be used to catalogue it. Seller must put so much emphasis as possible defining the object since it will be easier find a buyer. Moreover, the sellers choose the way the item will be offered, that is by means of auctions or directly sold to the first buyer.
- *Buyer:* He enters a specification of desired item and makes a search with the aim to get a list of items recommended. A specification consist of a set of characteristic, the values of that characteristic desired and the importance or relevance associated to each characteristic.

Due to the different users that interact with the portal and the different way to do it, any suggested portal design should consider the following points with regard to the use of characteristic and values:

- The administrator may define a new class of items. To do it, he should determine a new class, the characteristics that define to this class and the range of values that they could take.
- The administrator may modify one class previously defined, adding or removing characteristic and re-defining the range of values that these characteristics could take.
- The sellers may define the item they insert into the catalogue easily using non precise values. In addition to this they should take part in the definition of items classes and the characteristics and values that define them.
- The buyers may give specifications in which the values the characteristics take are given in a natural way for them, usually in a non precise way.
- The buyers may associate weights to the characteristics in order to increase or decrease its importance or relevance when a searching is carried out.
- The buyers may use values ranges of a characteristics when they make a search. The use of ranges increases the amount of objects closed with the object desired.

The suggested portal will include others important features of C2C e-commerce portals (ebay) such as:

- It allows to make lexical searches and by commercial models.

- It incorporates a message service to allow the communication between users. For example, it will be used when a user need to clarify different aspects of a particular item.
- It includes a register and manage users system which associates each nickname to a real user. This information will be used to allow to users get in touch or communication with.
- It allows to control items sale or auction time.
- It has a system of users votes with a double aim, on the one hand to associate to each user a respect of the community rate, on the other to know the prestige of each one of the users inside the community that uses the portal.
- It registers the sales of the portal with the aim to allow to the administrator use this information to improve the detail of some categories items.

We must point out here, that the specifications and descriptions are acquired directly from the users. In the next section, it will be explained how to allow this freedom when they communicates the characteristics and values.

3 CATALOGUING AND SEARCHING ITEMS

The main feature of the suggested method is that it gives license to the sellers when define an item and to the buyer when specify the characteristics of the item which they want to search. Therefore, they could use the following data types, which we have considered to be the most common (Castro-Schez et al., 2004b):

- Continuous or numerical, they can take values in the scale of real numbers (e.g. the house surface in m^2 or the price of a house).
- Graduated or ranking, they only can take values from a finite set of values (e.g. the number of rooms of a house or the house condition).
- Ordered-discrete or ordinal, they can take literal strings as value. However, these values can be arranged according to some judgement (e.g. the district where is placed a house, it can be arrange having into account where we prefer live).
- Unordered or nominal, they can take any sequence of words as value and it is not possible arrange them because there are infinite possibilities (e.g. the street in which is located a house).
- Boolean or logic, they can take only two values, i.e. *Yes* or *Not*. (e.g. if a house has connection to Internet or not).

We need a representation which allows all the previously mentioned values to be described. This representation must have the following properties:

- Mixed, it must allow to represent vague and precise values. The human mind presents this duality. The uncertainty could appear when a numerical characteristic is valued in an approximate way or several values are used to value a characteristic because its precise value is not known.
- Flexible, it must can hold any value used by the user (administrator, seller or buyer).
- Computable, it must can be manipulated by a computer program.

We will make use of the trapezoidal function used in the fuzzy logic field. This function has four parameters (a,b,c,d) and it allow us to represent the concrete and fuzzy values. The interval $[b,c]$ establishes the set of values that belong to the fuzzy value in a grade equal to 1. The values $[a,b)$ and $(c,d]$ establish the values that belong to the fuzzy value in a grade less than 1.

The parameters used to represent each of the types of characteristics before mentioned can be consulted in (Castro-Schez et al., 2004a).

Next, we must establish the method to insert and search items in the portal. Before explain how is carried out the insertion and search of a item into the catalogue, we need to define some previous concepts.

3.1 Previous Concepts

A catalogue, noted as $\mathcal{C}$, is defined as a set of classes of items, C_i.

$$\mathcal{C}=\{C_1, C_2, C_3,\ldots,C_n\}$$

A catalogue is a not static element. The classes that define it are not definitive and can be added, modified or removed from it.

A class consist of a set of items e_j, that is:

$$C_i=\{e_1, e_2, e_3,\ldots,e_k\}$$

Each class C_i is defined by a set of variables or characteristics V_{C_i} (from now we will refer to them as variables) so that:

$$V_{C_i}=\{v^1_{C_i}, v^2_{C_i}, v^3_{C_i},\ldots,v^m_{C_i}\}$$

The variables $v^j_{C_i}$ that define a class C_i must be representative of it, being of use in differentiating the items e_j that belong to the same class ($e_j \in C_i$).

Each variable $v^j_{C_i}$ has a definition domain where it can take values, we note it as $DDV^j_{C_i}$. This set, $DDV^j_{C_i}$, determines the values that can take an object e_j of the class C_i according to the variable $v^j_{C_i}$.

The definition set of a variable $v^j_{C_i}$, $DDV^j_{C_i}$ is defined as:

$$DDV^j_{C_i}=\{A_{j1}, A_{j2}, A_{j3},\ldots,A_{jk}\}$$

with A_{ji} being a fuzzy set defined by means of a trapezoidal function with parameters a,b,c and d. These parameters are established depending on the type of value that they represent.

The definition set of a class, noted as DDV_{C_i}, is defined as:

$$DDV_{c_i}=\{DDV^1_{c_i}, DDV^2_{c_i}, DDV^3_{C_i},\ldots,DDV^n_{c_i}\}$$

We can associated with the catalogue $\mathcal{C}$ the catalogue definition domain set, we noted it as $DDV_{\mathcal{C}}$, and we define it as the union of the definition domain sets DDV_{C_i} of each $C_i \in \mathcal{C}$, that is:

$$DDV_{\mathcal{C}} = \bigcup_{i=1}^{n} DDV_{C_i}$$

In this paper, it is not important the way of obtaining these sets, $\mathcal{C}$ and $DDV_{\mathcal{C}}$, though we show how could be acquired each C_i and DDV_{C_i}. We suggest make use of a semi-automatic method to acquire the variables V_{C_i}, together with the values the variables can take DDV_{C_i} from an expert human present in (Castro-Schez et al., 2004a). We may use this method for each $C_i \in \mathcal{C}$ and obtain $DDV_{\mathcal{C}}$.

3.2 Assessing Similarity

Initially, the seller or the buyer must give a detailed description of the item to be inserted into the catalogue or searched to be bought, respectively.

The description of such item, noted as o_i, consist of a set of features that it has or should have $(V = \{v_1, v_2, \ldots, v_k\})$ and the set of values C_{v_i} each feature v_i could take $(C_v = \{C_{v_1}, C_{v_2}, \ldots, C_{v_k}\})$.

Moreover, the variables that are used to describe the item o_i could have not the same importance for the seller or buyer who describes it. They has to indicate the importance of each variable v_i in V $(\forall v_i \in V)$.

Therefore, the description of the item o_i must associate to each variables v_i a importance P_{v_i}, that establish the importance that it has for the user that variable. Each P_{v_i} takes his value from the interval $[0,1]$). The greater the value P_{v_i} the greater the importance of the variable v_i. This allow to pay major attention to those variables more important for the user when the search is carried out (Castro-Schez et al., 2005).

To give more freedom to the user at the moment of making searches, the user can include in the descriptions of the item ranges of values. This option is not available for all the types of variables.

The use of ranges is allowed in continuous, graduated and ordinal variables. To give a range in continuous and graduated variables the user must insert two values, those that define the limits of the range. To use a range in ordinal variables implies to enumerate of set of literal that the user wants in his ideal object (e.g. the city districts set where the user would live).

The item o_i description quality will allow to carried out searches more or less precise and find items similar to o_i into the catalogue. In this way, the search can be adjusted in the catalogue to the real needs of the user.

We must search the items into the catalogue $\mathcal{C}$ that will be more similar to the item o_i. It implies to search into each class C_i of the catalogue $\mathcal{C}$ $(\forall C_i \in \mathcal{C})$.

The definition domain set of a class C_i, i.e. DDV_{C_i}, establishes a space of reference in which are placed all the items e_j that belong to the class $(e_j \in C_i)$ (see Fig. 1).

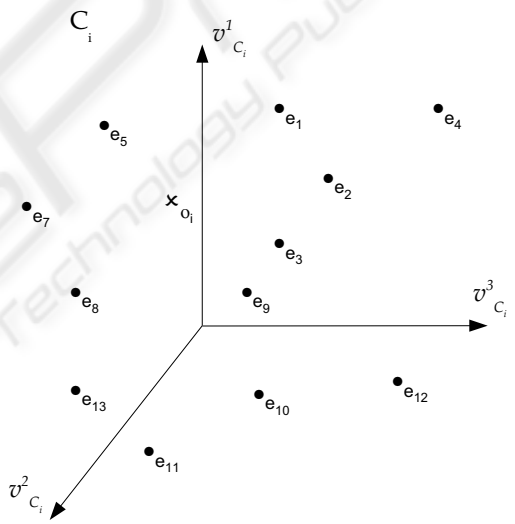

Figure 1: C_i space of reference.

To search into a class, C_i, implies to access to the space in which are defined the items belonging to this class (DDV_{C_i}) and study the proximity of o_i with the items e_j that are placed in this space $(e_j \in C_i)$. The items e_j closely located to the item o_i will be the candidates for being bought. On the other hand, if we are cataloguing the item o_i, we will make it in the class C_i whose items are closest to it.

As it can be observed, the search implies to study the distance between each item of the space, i.e. e_j, and the item o_i depending on the set of variables used for defining both items. Therefore, it is necessary to use a measure that provides the above mentioned information to us and that could be used in a uniform way without depend on the variable type.

Since all values that a item takes depending on a variable v_i are represented as trapezoidal numbers, the local distance between two items e_j and o_i according to a variable v_i could be assessed by means of a measurement based on the calculation of the area existing between the two fuzzy values, A and B that take these items, e_j and o_i, respectively, in the variable v_i. Such measurement and its properties was suggested in (Castro-Schez et al., 2004a) and we refer to it as $d(e_j, o_i, v_i)$. In this paper, we use the normalized measurement of distance $d_N(e_j, o_i, v_i)$ (Castro-Schez et al., 2004a), however, any measure d that establishes the proximity or separation between fuzzy sets represented by means of the above mentioned functions could be used.

The global distance between two items e_j and o_i is calculated as sum of the distances local between those items according to the set of variables used to define both items. That is, if $e_j \in C_i$ and o_i is described by means of set of variables that we call V, then we must have into account the set of variables $V \cap V_{C_i}$. In this point, we suppose that both sets have the same terminology. If this was not the case, a method must be study to obtain it.

The global distance between two objects e_j (with $e_j \in C_i$) and o_i, noted as $D(e_j, o_i)$, will be calculated as:

$$D(e_j, o_i) = \sum_{\forall v_i \in V \cap V_{C_i}} d_N(e_j, o_i, v_i) \times P_{v_i} \qquad (1)$$

The importance associated to each variable P_{v_i} is considered in this calculation.

Once the distance between o_i with each item $e_j \in C_i$, $D(e_j, o_i)$, has been calculated for all classes of the catalogue ($\forall C_i \in \mathcal{C}$), the system returns a list arranged by the distance value with a set of items closest to o_i. The recommended item will be the one with the lowest distance value.

The searching process in pseudocode is the following:

- Input: $\mathcal{C}$, o_i ($V = \{v_1, v_2, \ldots, v_k\}$, $C_{v_i} = \{C_{v_1}, C_{v_2}, \ldots, C_{v_k}\}$ and $P_V = \{P_{v_1}, P_{v_2}, \ldots, P_{v_k}\}$).
- Output: $\mathcal{L} = \{e_i, \ldots, e_j\}$ | $e_i \in C_i$ and $C_i \in \mathcal{C}$.

1. For each $C_i \in \mathcal{C}$ do
   (a) For each $e_j \in C_i$,
      i. Calculate $D(e_j, o_i)$,
2. Arrange e_j according to the value $D(e_j, o_i)$ in $\mathcal{L}$.
3. Recommend the object e_j with the lowest value $D(e_j, o_i)$.

To make more efficient the search we could determine a subset of classes $\mathcal{D}$ ($\mathcal{D} \subsetneq \mathcal{C}$) in which to carry out the search. The set $\mathcal{D}$ will consist of classes C_i belonging to $\mathcal{C}$ that possess a number of variables commons to the specification o_i great than mean (τ). That is,

$$|V_{C_i} \cap V| \geq \tau$$

Moreover, the user could also select the class into the which he want to search, C_i. In this case, the user should use the variables that define to that class V_{C_i}.

4 AN APPLICATION EXAMPLE

This section show briefly the application of the search method suggested to an example.

The user selects the class in which is interested, *rent houses*, and studies the variables that define to this class with the aim to value his *ideal item*, o_i, according to this information. For this case, the set of variables that are used are shown in Table 1.

Table 1: Variables that define rent houses class.

Variable	*Type*	*Value range*	Domain
Price	Continuous	Yes	$[100, 700]$
Rooms	Graduated	Yes	$[1, 5]$
District	Ordinal	Yes	$[1, 4]$
Internet	Boolean	No	
Street	Nominal	No	

The definition domain of the *Price* variable is shown in Fig. 2.

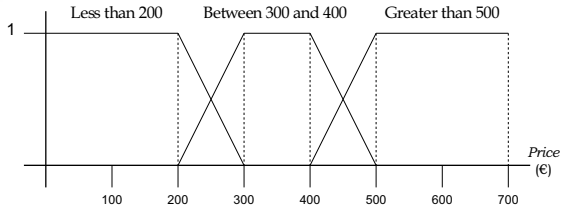

Figure 2: Definition of the *Price.* variable.

The user values each variable and associates to each one a importance grade (see Table 2).

Table 2: Description of the user item, o_i.

Variable	*Value*	*Range*	*Importance*
Price	(300,400,400,500)	No	Maxim (1)
Rooms	3	No	High (0,75)
District	$[1, 3]$	Yes	Low (0,5)
Internet	Yes	-	High (0,75)
Street	Altagracia	-	Low (0,5)

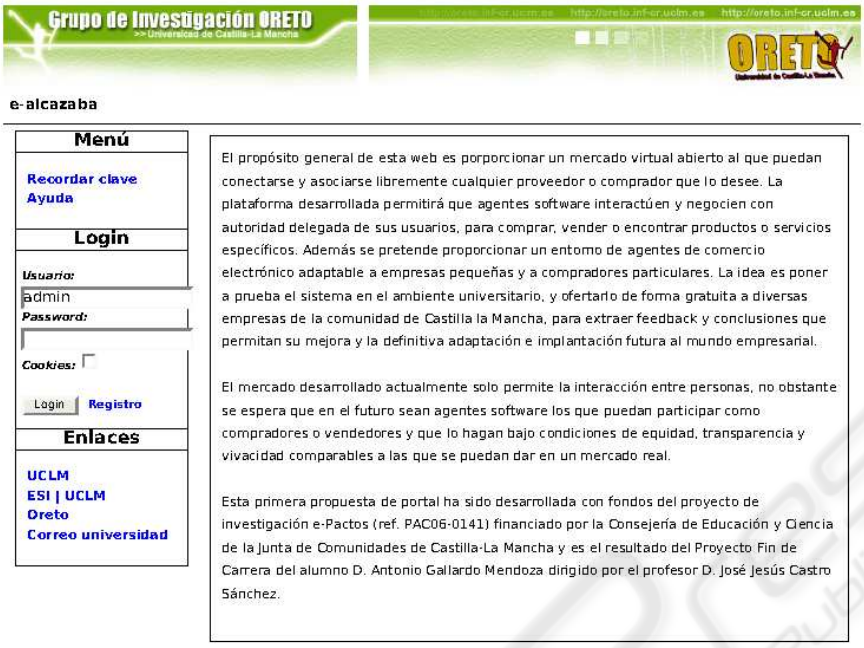

Figure 3: Portal's screen capture.

Table 3: Items of rent houses class.

House C_i	*Price* $v^1_{C_i}$	*Rooms* $v^2_{C_i}$	*District* $v^3_{C_i}$	*Inter.* $v^4_{C_i}$	*Street* $v^5_{C_i}$
e_1	*550* $a = 500$ $b = 550$ $c = 550$ $d = 600$	5	4	Yes	De la Rosa
e_2	200	2	4	No	Altagracia
e_3	300	4	2	No	De la Luz
e_4	*350* $a = 300$ $b = 350$ $c = 350$ $d = 400$	3	3	Yes	Pilar
e_5	500	4	1	Yes	Postas

Table 4: Local distances $d(e_j, o_i, v_{C_i})$.

House	*Price*	*Rooms*	*District*	*Inter.*	*Street*
e_1	0,13	0,5	0,25	0	1
e_2	0,25	0,25	0,25	1	0
e_3	0,08	0,25	0	1	1
e_4	0,01	0	0	0	1
e_5	0,08	0,25	0	0	1

After, the user has given this information that define his ideal item (Table 2), the system proceeds to search similar items into the class (*rental houses*).

If we suppose that the items which are stored in the portal under the class C_i (*rent houses*) are those that show in Table 3. The values in cursive are approximate values, and we show how they are defined by means of a fuzzy number.

The local distance between the item o_i and each item e_j from the class C_i ($e_j \in C_i$), i.e. $d_N(e_j, o_i, v^k_{C_i})$ for $k = 1$ to 5 are shown in Table 4.

Next, we calculate the global distance between each e_j and o_i, applying Eq. 1. The following ordered list is obtained:

$$e_4(0,51), e_5(0,77), e_1(1,14), e_2(1,32), e_3(1,52)$$

The application recommends to rent the item e_4 since it has the lowest global distance value $(0,51)$.

5 CONCLUSIONS

In this paper, we have suggested a new method based on fuzzy logic to catalogue and search items (goods or services) in a C2C e-commerce portals. It allow to the buyer to carry out searches in an imprecise way. The seller may insert items into the catalogue in fuzzy terms.

Recently, agent technologies are being applied to e-commerce where a personalized, continuously running, semi-autonomous behavior is desirable (Guttman et al., 1998; He et al., 2003). The agents could be integrated easily in this portal. It will be our main research for the future.

The prototype of the suggested C2C e-

commerce portal based on fuzzy logic has been put at the following web address (see Fig. 3): `http://apps.oreto.inf-cr.uclm.es/e-alcazaba/`

ACKNOWLEDGEMENTS

This work has been supported by Research Projects e-PACTOS (ref. PAC-06-141) and SARASVATI (ref. PBC06-0064) founded by Junta de Comunidades de Castilla-La Mancha (JCCM).

REFERENCES

Cameron, D. (1997). *The New Business Platform for the Internet*. Computer Technology Research Corporation, Charleston, 1st edition.

Castro-Schez, J. J., Castro, J. L., and Zurita, J. M. (2004a). Fuzzy repertory table: A method for acquiring knowledge about input variables to machine learning algorithm. *IEEE Transactions on fuzzy systems*, 12(1):123–139.

Castro-Schez, J. J., Jennings, N. R., Luo, X., and Shadbolt, N. R. (2004b). Acquiring domain knowledge for negotiating agents: a case of study. *International Journal of Human-Computer Studies*, 61(1):3–31.

Castro-Schez, J. J., Jimenez, L., Moreno, J., and Rodríguez, L. (2005). Using fuzzy repertory table-based technique for decision support. *Decision Support Systems*, 39(3):293–307.

Chaudhury, A. and Kuilboer, J. P. (2001). *E-Business and E-Commerce Infrastructure: Technologies Supporting the E-Business Initiative*. McGraw-Hill, Boston, 1st edition.

Compras. Compras web site http://www.compras.com.

Ebay. Ebay web site http://www.ebay.com.

Froogle. Froogle web site froogle.google.com.

Guttman, R., Mouka, A., and Maes, P. (1998). Agent-mediated electronic commerce: A survey. *Knowledge Engineering Review*, 13(2):147–159.

He, M., Jennings, N. R., and Leung, H. (2003). On agent-mediated electronic commerce. *IEEE Transactions on Knowledge and Data Engineering*, 15(6):1345–1363.

Kowtha, N. R. and Choon, T. W. I. (2001). Determinants of website development: a study of electronic commerce in singapore. *Information & Management*, 39(3):227–242.

Laudon, K. C. and Laudon, J. P. (2005). *Management information systems: managing the digital firm*. Prentice Hall, NJ, U.S.A, 9th edition.

Turban, E., Lee, J., King, D., and Chung, H. (2000). *Electronic Commerce: A Managerial Perspective*. Prentice-Hall, Upper Saddle River, NJ, 1st edition.