A Web-based CBR System for e-Mail Response

Young Suk Yoon[a], Jae Kwang Lee[b], Chang Hee Han[c]

[a]*Graduate School of Management, Korea Advanced Institute of Science and Technology*
207-43, Cheongryangri-Dong, Dongdaemun-Gu, Seoul, 130-012, Korea
Tel: +82-2-958-3684, Fax: 82-2-958-3675, e-mail: ysyoon@kcals.or.kr

[b]*Department of e-Business, Korea Polytechnic University*
2121, Jungwang-dong, Shihung, Kyunggi-do, 429-793, Korea
Tel: +82-31-496-8424, Fax: 82-31-496-8309, e-mail: jklee@kpu.ac.kr

[c]*Department of Business Administration, Hanyang University*
1271, Sa-1-Dong, Sangrok-Gu, Ansan, Kyunggi- Do, 425-791, Korea.
Tel:+ 82-31-400-5634, Fax: 82-31-400-5591, e-mail: chan@hanyang.ac.kr

Abstract

Due to the rapid growth of Internet, means of communication with customers in a traditional customer support environment such as telephone calls are being replaced by mainly e-mail in a Web-based customer support system. Although such a Web-based support is efficient and promises potential benefits for firms, including reduced transaction costs, reduced time, and high quality of support, there are some difficulties associated with responding to many types of customer's inbound e-mails appropriately. As many types of e-mail are received, considerable attention is being paid to methods for increasing the efficiency of managing and responding e-mails. This research proposes an intelligent system for managing customer's inbound e-mails in organizations by applying case based reasoning technique for responding to various customers' inbound e-mails more effectively. In this approach, a case is represented as a frame-typed data structure corresponding to an inbound e-mail, keywords, and its reply e-mail. In the retrieval procedure, keywords and affinity set is developed to index a case, and then the case is represented as a vector, a case vector. Also, cosines value is calculated to measure the similarity between a new inbound e-mail and the cases in the case base. In the adaptation procedure, we provide several adaptation strategies to adapt and modify the retrieved case. The strategies guide to make an outbound e-mail using product databases, databases for customer support, etc. Additionally, the Web-based system architecture is proposed to implement our methodology. The proposed methodology and system will be helpful for developing more efficient Web-based customer support.

Keywords: customer support, e-mail response, case based reasoning

Introduction

Recently, customer support or service in organizations is one of the most important business improvement theme in improving their business competences. Furthermore, many firms have realized, as their marketplaces have become more global and service oriented, that customer support is critical to their competitiveness (Negash et al., 2002). Customer support is an important activity in all types of industries. The importance of such support facilities cannot be overemphasized. Many customers view customer support as one of the most important criteria when evaluating a product or a service (Foo et al., 2000). Also, customer service has a strong link to customer satisfaction, which then yields customer loyalty and long-term profitability (Szymanski & Henard, 2001; McKenna, 1991).

In a traditional customer support environment, mainly call centers or service centers are responsible for receiving inquiries from their customers via telephone calls. Due to the rapid growth of the Internet with its widespread acceptance and accessibility, means of communication with customers in the traditional customer support center, such as telephones, letters, and direct-visiting, are being replaced by e-mails and bulletin board systems (BBSs) using the Internet constantly (Choi et al., 2003). The web, e-mail and chat are new electronic customer service – eService – tools for enhancing customer relationships (Barnes & Cumby 2002; Zemke & Connellan, 2001; Newell, 2000; Reichheld & Shefter, 2000). Communications between a firm and its customers, other than face-to-face discussions, take place through one or more media, via interactions with the media by both parties. The features of a web-based interface make it an attractive choice as a medium for interaction between the firm and its customers (Hoffman & Novak, 1997).

It is clear that web-based customer support systems are

important (Negash, 2002). However, there are also some difficulties associated with responding to many types of customer inquiries appropriately in spite of all the above benefits. In this research, the original project was prompted by a local large-sized company in Korea, called 'K' company, for its customer service needs. In order to provide an appropriate response to their customer inquiries, the means of communication in their web-based customer support system can be classified the following: e-mails, a list of frequently asked questions (FAQs), BBSs, and direct-searching. In case of 'K' company, customers use e-mail, as a mean of communication, almost 70% rather than all the other means such as direct searching, FAQs, and BBSs in order to have an appropriate answer by themselves. In this fact, we feel sure that customer support with using e-mail becomes more general method to communicate its customers in web-based customer support system.

As the large amount of e-mail received, there is now substantial attention being paid to methods for increasing the efficiency of processing and organizing messages (Camino et al., 1998). E-mail can enhance customer relationships, customer satisfaction, and profitability (Murphy & Tan, 2002). However, it is difficult to classify many different types of incoming e-mail from many different types of its customers in the web-based customer support system. Especially, a large-sized company which has several divisions and deals with many products and services is always distressed by this kind of problems. For example, the 'K' company received many different types of incoming e-mail such as recruiting, products and services information, IR information, and so on through one single channel in the system. Because of the difficulty of classifying inbound e-mails into appropriate respondents or divisions, the company has problems which are the followings. The response process is time-consuming and costly. It also needs to keep on training new service people, and at the same time, come up with new incentive scheme to keep experienced service people. The database of service records is only used by the service people. Expert advice to the problem is given either through the experience of the service people or the available past service information in the service database (Foo et al., 2000).

In this research, we propose an intelligent methodology and system for managing customer's inbound e-mails in organizations by applying case-based reasoning (CBR) technique to responding various customers' inbound e-mails more effectively. The basic idea of CBR is that humans reuse the problem solving experience to solve a new problem (Kolodner, 1991). This proposed approach regards a reply e-mail of one inbound e-mail as a case, so it stores replies of similar e-mails at a case base. Knowledge acquisition is usually known to be a most difficult and time-consuming process in knowledge-based systems. However, CBR can acquire knowledge with ease using inductive methodology, so it is useful especially when knowledge is incomplete, or evidence is sparse (Kolodner, 1993). In addition, CBR works well in domains that are poorly understood, because the system does not need to know why something worked in the past (Kolodner, 1991; Lee & Kim, 2002).

The main task of using CBR is generally the representation of a case, a retrieval procedure, and an adaptation procedure (Kolodner, 1991; Kolodner, 1993; Lee et. al., 2000). This proposed approach represents a case as a frame-typed data structure corresponding to an inbound e-mail, keywords, and its reply e-mail. In the retrieval procedure, keywords and affinity set is developed to index a case, and then the case is represented as a vector, a case vector. Also, cosines value is calculated to measure the similarity between a new inbound e-mail and the cases in the case base. In the adaptation procedure, we provide several adaptation strategies to adapt and modify the retrieved case. The results showed that our approach for e-mail response provides customer managers with robust knowledge-based support.

2. Web-based CBR system for e-mail response

2.1 The procedure of the system

In this section, we briefly illustrate web-based CBR system. Figure 1 shows the procedure of CBR system which has three major phases such as case representation, case retrieval, and case adaptation.

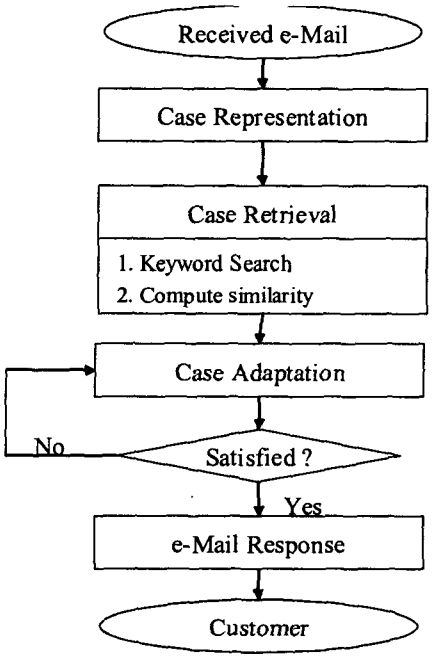

Figure 1. The procedure of the system

2.1.1 Case representation and keyword search

To represent an e-mail as a case, a frame typed case is defined as three components, an *inbound e-mail*, *keywords*, and its *reply e-mail*. If an e-mail from customer is received, the mail is to be represented as a frame-typed case representation. Initially, the e-mail can be translated as a frame that has only one component, *inbound e-mail*. After text searching the e-mail for deriving keyword set, the *keywords* of frame-typed case will be established. Then the

reply e-mail will be constructed after the response e-mail is established. Thus, the form of a case is established by the completion of the procedure from the *inbound e-mail* up to the *reply e-mail*.

2.1.2 *Case retrieval*

The degree of similarity between a new inbound e-mail and cases in the case base is calculated to retrieve the most similar case from the case base. The similarity is a case vector that is derived from keyword and affinity network. The keyword affinity network represents the synonyms of keywords as a network form. Also, relations of synonyms are quantified in the network. And then, under the consideration of the synonym in the network, *keywords*, which is a form of a vector, is converted into case vector. Using the quantified keyword affinity network can increase the accuracy in calculating the similarity.

2.1.3 *Case adaptation*

Generally, the most similar case from case retrieval phase has to be modified and updated for better result or response to a newly received *inbound e-mail*. Several adaptation strategies are introduced to make a reply e-mail in section 4. In this phase, the files which are already existing and utilized for customer support or product development are attached for the completion of a *reply e-mail*. Now, since *reply e-mail*, a component of a case, is established for a new case, this new case that has three components of case representation can be stored in the case base. And then, the CBR system translates the case into a e-mail form, and sends the reply e-mail to the customer.

2.2 Web based CBR system architecture

For the development of the system, as indicated in the architecture in figure 2, we propose a composition of four layers; interface layer, case handler layer, manager layer and storage layer. In *storage layer*, keyword affinity set, case base, and attachment file set are separately stored in indivisible databases. The database for keyword affinity set stores the keywords and their synonyms as a network form. The attachment file set stores several product files and customer support documents, etc. The case base contains a lot of frame typed cases which are already translated into reply e-mails and sent to customers. In *manager layer*, three modules; keyword manager, case manager, and data manager are needed to manage each database or set in storage layer.

In *case handler layer*, four major calculation and search modules are included. Case value calculator module computes each component values of case vector. Similarity calculator derives the similarity value between a new inbound e-mail and cases in case base, and determines the most similar case. Text search modules find keywords in a new customer's inbound e-mail and assign the keywords to *keywords* of frame typed case. Case editor is used for revising and updating the cases in case adaptation phase. In *Interface layer*, case and e-mail translator translates an e-mail into a frame typed case form, and vice versa.

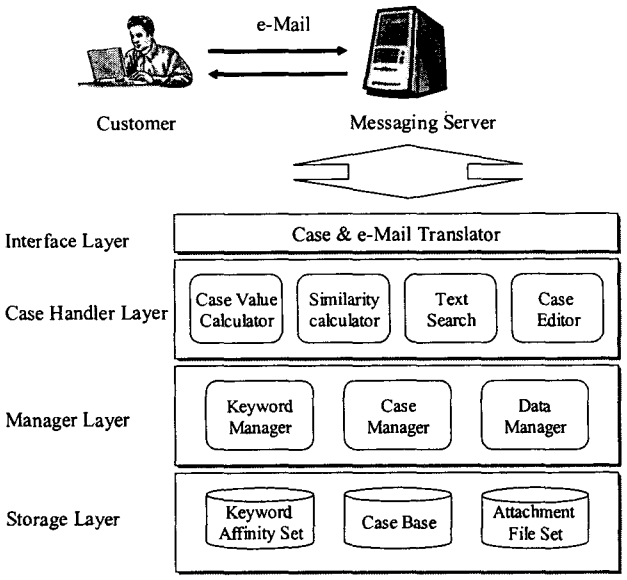

Figure 2. The system architecture

3. Case representation and case retrieval

3.1 Case representation

In this research, we use a frame to represent the case. A case is composed of an inbound e-mail, keywords, and its reply e-mail. A case of the case base contains all the 'INBOUND E-MAIL' and 'REPLY E-MAIL' related information such as product category, location, sender, date, contents, attached file, and keywords of the case. We focused on replying the e-mail response for the inbound e-mail, so it does not contain any administrative information.

INBOUND E-MAIL
PRODUCT CATEGORY 1 : Computers & Related
PRODUCT CATEGORY 2 : Printer
LOCATION : PUSAN KOREA
SENDER : Simon Peter
DATE : 20030725
SUBJECT : Printer Driver
CONTENTS : 2개월 전에 K company에서 생산한 프린트를 구매하여 사용하고 있습니다. 그런데, 제가 최근에 PC를 교체하여 프린트를 연결하였으나 프린트가 정상적으로 작동을 하지 않습니다. 도움을 바랍니다. 참고로 제가 사용하고 있는 프린트의 모델 명은 *ML2100*입니다.

KEYWORDS
(K1 : 1) (K4 : 1) (K5 : 1)

REPLY E-MAIL
SENDER : S.M Kim
DATE : 20030726
SUBJECT : Printer Driver
CONTENTS : 보내주신 이메일 잘 받았습니다. 문의 하신 내용이 구적이지 않아 정확한 원인을 알 수 없으나, PC를 교체하신 후 프린트 모델에 적합한 드라이브를 인스톨하지 않은 것 같습니다. 일단, ML2100 모델에 적합한 드라이브를 첨부하여 보냅니다. 이 드라이브를 설치하신 후에도 정상적으로 작동이 되지 않으면, 가까운 고객지원 센터를 방문하여 주시면 감사하겠습니다. 고객님의 근처에 있는 폐사의 고객지원 센터에 관한 정보(위치, 담당자, 연락처 등)를 아래에 첨부하였습니다. 감사합니다.
Customer manager, S.M Kim
ATTACHED : ML2100.DRV, CS_center.doc

Figure 3. An example of the case representation

. The 'KEYWORDS' about 'INBOUND E-MAIL' does a very significant role in our approach. When retrieving a case from the case base, one of important criteria is the degree of

similarity between a new e-mail and the case of case base. The frame-typed representation of an example case is shown like Figure 3. The case contains the information about inbound e-mail, keywords, and its reply e-mail. The 'INBOUND E-MAIL' information contains product category, location, sender, date, subject, and contents. The 'REPLY E-MAIL' information contains has information contains sender, date, subject, contents, and attached file. Also, the case contains keywords for case retrieving procedure. Each keyword has a name and its value such as (K1:1). The keyword values are an integer value, such as 1 or 2. It means the frequency of the keyword in the inbound e-mail. As an example, Figure 4. shows the hierarchical structure of case example.

3.2 Similarity-based case retrieval

3.2.1 Case indexing using keywords and its affinity

In this research, assumed m cases are stored in a case base. We can represent the case as keyword vectors of the form

$$E_k = (a_{k1}, a_{k2}, \ldots\ldots, a_{kn})$$

, where the coefficient a_{ki} represents the value of keyword i in case E_k. Typically a_{ki} is to be an integer number when keyword i appears in case E_k, and 0 when keyword i is absent in case E_k.. Following shows an example of initial case vector, where the case is represented by nine keywords.

$$E_1 = (1, 0, 0, 1, 1, 0, 0, 0, 0)$$

However, keywords alone are not enough to represent the case, so this research uses synonym to represent the case exactly. The affinity value among synonyms has a number between 0 and 1, and it is stored at a network-type knowledge base. For example, the "customer" and "consumer" may be used as a synonym, and the affinity exists between the two words. The keyword affinity network for the above example is assumed to be stored at knowledge base in advance as the following Figure 5.

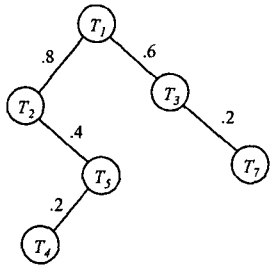

Figure 5. An example of keyword affinity network

The case vector is generated based on the initial case vector supplemented from the keyword affinity of the keyword affinity network. In that case, it is computed *an affinity value* among all keywords in index, which represents the degree of similarity. If a directed link between keywords does not exit in keyword network, the affinity value between keywords, T_i and T_j is computed by the following equation;

$$Affinity(T_i, T_j) = Max\{Min[Affinity(T_i, T_k), Affinity(T_k, T_j)]\}$$

$$, k = 1, \cdots, n.$$

The affinity values between all keywords in index are computed. If $Affinity(T_i, T_j)$ is not zero between keywords T_i and T_j, where a_{ki} (the value of keyword i in case E_k) is zero, and a_{kj} is not zero, then a_{ki} is replaced by the product of $Affinity(T_i, T_j)$ and value of a_{kj}. For example, assume that keyword affinity network is given like Figure 5, where the affinity value between keyword T_2 and T_1 is 0.8. At first, only keywords T_1, T_4, and T_5 are assumed to be appeared in E_1, but considering the keyword affinity, E_1 becomes related with keywords T_2, T_3, and T_7 also. It will be found that a_{12} is replaced with 0.8, $Affinity(T_1, T_2)$. Where the element of case vector E_k is 0, to reflect keyword affinity, the value of the element is computed from initial case vector and keyword affinity network. Following shows the case vector reflecting keyword affinity relations.

$$E_1 = (1, .8, .6, 1, 1, 0, .2, 0, 0)$$

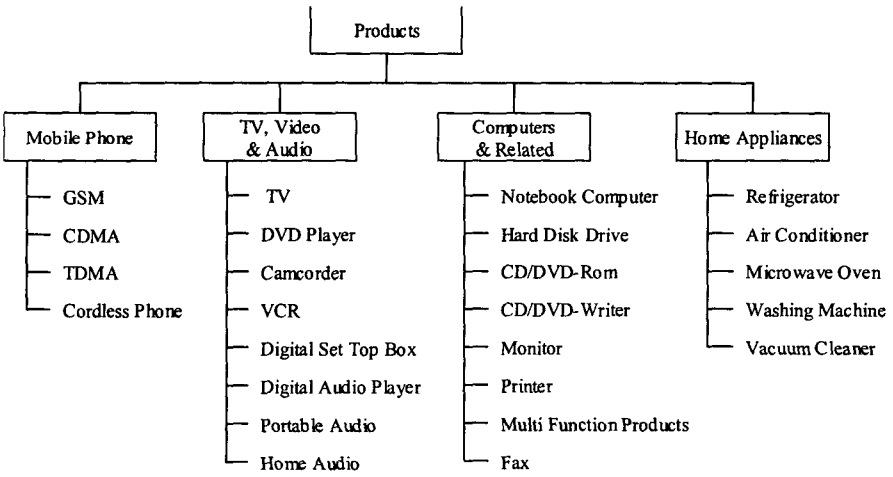

Figure 4. An example of the case representation

3.2.2 Similarity

To retrieve the best case, we calculate the degree of similarity between the new e-mail and cases in the case base. The basic idea is that if new e-mail and stored cases have similar keywords and their frequencies are also similar, then it is concluded that they are similar. The similarity degree between new e-mail and stored cases are calculated from the case vectors. In this research, one case is represented by n-dimensional vector, E_i, where $i = 1,...m$, and m is the number of case. Therefore, the similarity S_i between new e-mail E_0 and the stored case E_i is computed as follows;

$$S_i = \|Sim(E_0, E_i)\| ,$$

where $Sim(E_0, E_i) = \frac{E_0 \cdot E_i}{|E_0| \cdot |E_i|} = \cos\theta \quad 0 \le \theta \le \frac{p}{2}$.

Hence, the *similarity degree* or *similarity value* between new e-mail and stored cases are represented by the cosine value of vector E_0 and vector E_j. Following presents the similarity between new e-mail E_0 and the stored case E_1.

$S_1 = Sim(E_0, E_1) = 3.4 / 5.32 = .64$,
where $E_0 = (0, 2, 1, 0, 1, 0,1, 0, 0)$ and
$E_1 = (1, .8, .6, 1, 1, 0, .2, 0, 0)$.

4. Knowledge-based case adaptation

A case adaptation procedure supports to establish reply e-mail using the retrieved case. Three kinds of knowledge are necessary for e-mail response. The adaptation procedure relies on *domain-specific knowledge*, *systemic knowledge on the company code of customer support,* and *adaptation strategies*. When making the reply e-mail, the respondent bases the reply on the retrieved case and also on domain-specific knowledge, systemic knowledge on the company code of customer support, and adaptation strategies that the system provides. Domain-specific knowledge is knowledge on the inquired product and is provided in a form of product information database. Systemic knowledge on the company code of customer support is provided in a form of operation manual. Also, adaptation strategies provide a guideline on how to use these two kinds of knowledge; domain-specific knowledge and systemic knowledge.

In this system, an adaptation procedure starts from a case and modifies it considering the reply e-mail until the modified reply e-mail is suitable to the inbound e-mail. Modification process starts from the difference between inbound e-mail of past case and new inbound e-mail. Then a new reply e-mail is created using domain-specific knowledge, an appropriate tool for the differences between the past case and the new case. When a new reply e-mail is created, it is checked whether it is appropriate to the company code of customer support and when it is found satisfactory, the reply e-mail is sent. This e-mail is then stored in the case base.

The adaptation procedure is presented as follows:

Step(1) Discovering the difference between past case and new case.

In this step, the locations, data, subject, contents and keywords of the past inbound e-mail and the new inbound e-mail are checked for differences.

Step(2) Modifying the reply e-mail of the past case.

To match past reply e-mail with the new inbound e-mail, domain-specific knowledge and adaptation strategies can be used. The modification process modifies the 'CONTENTS' and 'ATTACHED' which are included in past reply e-mail. This step is repeated until the differences are solved.

Step(3) Check the company code of customer support.

For the verification of the company code step, systemic knowledge about the company code of customer support is used to check and to correct whether the reply e-mail is satisfactory with the company code. If the ID does not satisfy the company code then the reply e-mail is modified again.

Step(4) Updating case base.

Finally, the resulting reply e-mail, inbound e-mail and the relevant keywords are stored to the case base as a new case.

5. Conclusion

This paper has presented a new way for customer support and responses using CBR approach. In order to develop more efficient web-based customer support, systematic approach to customer support management using CBR technique has been rarely seen in literature. Therefore, this paper suggests a new approach to customer support management by using CBR approach. Research on customer support and web-based e-mail response systems will increase rapidly due to the widespread of Internet technology. Although various issues have been raised in web-based customer support system, since the customer support via e-mail is increasing rapidly, more widespread researches in the related field of web-based CBR system for e-mail response are expected.

Furthermore, with the application of the proposed system in our study, a tremendous cut in customer service cost, decrease of repeated customer's inquiries, establishment of knowledge-base for corporate customer's needs and response, efficient management and supervision of customer's inquiries, reduction of customer's complaints by quick and accurate response, improvement of service through adaptation of customer's views are expected.

Our future study would be not only to find an improvement through demonstrative effective study application of this research, but also improvement of the system to be adaptable all through-out the web-based customer support system and customer support through e-mail.

References

[1] Barnes, J. G and J.A. Cumby (2002) Establishing customer relationships on the Internet requires more than technology. *Australasian Marketing Journal*, 10(1), 36-36.

[2] Camino, B.M., A.E. Milewski, D.R. Millen and T.M. Smith (1998) Replying to email with structured responses. *International Journal of Human and Computer Studies*, 48, 763-776.

[3] Choi, H.R, K.R. Ryu, J. Kang, J.I. Shin and C.S. Lee (2003) An Automatic Question Routing System using Machine Learning. *Proceedings of Korea Intelligence Information Systems*.

[4] Foo, S., S.C. Hui, P.C. Leong and S. Liu (2000) An integrated help desk support for customer services over the World Wide Web – a case study. *Computers in Industry*, 41, 129-145.

[5] Hoffman, D.L. and T.P. Novak (1997) A new marketing paradigm for electronic commerce, *Information Society* 13 (1), 43-54.

[6] Kolodner, J. (1991) Improving Human Decision Making through Case-Based Decision Aiding. *AI MAGAZINE*, Summer.

[7] Kolodner, J. (1993) *Case-Based Reasoning*. CA: Morgan Kaufmann Publishers.

[8] Lee, J.K., Kim, J.K. & Kim, S.H. (2000) A Case-Based Reasoning Approach to Build an Influence Diagram. Intelligent Systems in Finance, Accounting and Management, 9(1), 55-63.

[9] Lee, J.K. and J.K. Kim, (2002) Case-Based Decision Class Analysis for Building a Decision Model, Expert Systems, 19(3), 123-135.

[10 McKenna, R. (1991) Relationship marketing. New York: Addison-Wesley.

[11] Murphy, J. and I. Tan (2002) Journey to nowhere? E-mail customer service by travel agents in Singapore. *Tourism Management*.

[12] Negash, S., T. Ryan and M. Igbaria (2002) Quality and effectiveness in Web-based customer support systems. *Information and Management*, 2029, 1-12.

[13] Newell, F. (2000) Loyalty.com: *Customer relationship management in the new era of Internet marketing*. New York: McGraw Hill Professional Book.

[14] Reichheld, F. and P. Shefter (2000) E-loyalty : Your secret weapon on the web. *Harvard Business Review*, 78(4), 105-114.

[15] Szymanski, D.M. and D.H. Henard (2001) Customer satisfaction : A meta-analysis of the empirical evidence. *Journal of the Academy of Marketing Science*, 29(1), 16-35.

[16] Zemke, R. and T. Connellan (2001) E-service. New York: Amacom.