Creation of a Customer-Communication Knowledge Web for High Performance Organization

Tapati Bandopadhyay

ICFAI Business School
IDPL Complex,Old Delhi Jaipur Road,Gurgaon-122016,India
91-124-2455485, Fax: 91-124-2455484
tapati@ibsdel.org,

Pradeep Kumar

ICFAI Business Schoo
IDPL Complex,Old Delhi Jaipur Road,Gurgaon-122016,India
91-124-2455485, Fax: 91-124-2455484
pkgarg@ibsdel.org

ABSTRACT

One of the most common characteristics of High Performance Organizations (HPO) is their highly customer-focused purposes. The agility which makes HPOs stand apart in the maddening crowd of competition lies in their ability to foresee or apprehend, and consequently, adapt to major changes in the environment, of which the market and customer demands are the major elements for consideration. Customer communication, therefore, is a precious knowledge asset for HPOs. In this paper, we are proposing a method to create a knowledge web using customer communication. We have integrated the primary communication elements as messages and then have developed the method based on similarity measurements to create a high-dimensional network representing various messages of cluster of similar messages as nodes. This network being high-dimensional is being referred in this paper as a Knowledge Web, which can facilitate an HPO to utilize the customer communications in a semantically effective way and also to integrate it to various analytical systems supporting decision making.

INTRODUCTION

In High Performance Organizations(HPO), competitive performance has been shown (Holbeche 2005) to be linked to as firm's ability to adapt to major changes in the environment. The environment includes the supply chains, political, economic, social and industrial environments, technological scenarios (as collectively called PEST or STEP in the field of Strategic Management), and, essentially, the customers with a range of requirements which are only dynamically stable. Customers, to High Performance Organizations, are treated as one of the most valuable sources of knowledge and the primary motivation for innovations. These organizations try to continuously learn from various sources including customer interactions. (Rowden 2001) argues that in unpredictable and highly turbulent business conditions, an organization's capacity to learn as it goes may be the only true source of competitive advantage. Learning,

therefore, is not only a source of competitive advantage; it has become an absolute necessity for survival.

Few characteristics of High performance Organizations (Collins 2001) show that:

1. They have highly customer-focused purposes.
2. They value employees.
3. They look for sustainable success over long-term

Based on these characteristics, an HPO model has the following key elements:

1. Readiness for change with constant learning, flexibility and speed
2. creation of a knowledge-rich context for innovation
3. boundary-less organizations
4. value-based management

Among these key elements, creation of a knowledge-rich context for innovation is the element that is in focus in the context of this paper. This is concerned with constant development and improvement of new and existing practices respectively, to nurture an organizational culture that is conducive to break-through ideas as well as concepts for continuous improvement and quality practices like TQM and Kaizen etc. Dealing with innovations and break-through ideas, there is almost always a high degree of failure, risk, uncertainty and complexity. Various approaches of Knowledge Management (KM) and Learning Organizations (LO) concepts aim towards reducing these elements. The primary requirements for a knowledge-rich context conducive to innovation can be broadly identified as:

- developing best practices which are self-evolving so that continuous learning and improvement is promoted, facilitated and encouraged successively
- promoting the Knowledge-Sharing culture as part of continuous learning in an organization
- maximizing the potential value of shared knowledge
- managing diversity
- managing change.

Creation of a Knowledge-Rich context for a High Performance Learning Organization is a challenge in itself. It encompasses almost all conceivable dimensions of an organization i.e. the HR perspectives (e.g. in terms of linking individual's performance to knowledge sharing or knowledge creation and group-knowledge creation), operations perspectives (e.g. feasibility analysis for a new idea with uncertain outcomes or usability predictions) and, implementation-wise, the technology dimensions that is how to extract knowledge from various resources, how to represent them in an integrated, usable form, and most importantly how to use them. For example, most of the high-performance companies have established systems for collecting customer inputs and feedbacks in various structured (e.g. forms/field based or GUI combo-box-based forms), semi-structured (e.g. forms with text-box inputs) or unstructured (e.g. e-mail messages) forms. Now, these can serve as a collection of valuable assets for knowledge creation in the

organization. But how this asset will be converted in a cost-effective, less resource-hungry and optimally usable form, is a real challenge.

VALUE OF CUSTOMER COMMUNICATION TO HPOS

Successful organizations treat customers as Gods, as the only way of their survival as well as gaining competitive advantage. They focus intensely on customers and their needs. They invest in systems like CRM (Customer Relationship Management), processes and often in capturing inputs also to improve their products and provide better services. They focus on retaining customer loyalty as well as on attracting new customers. In Roffey Park research (2004), high performance has been shown to be strongly correlated to customer-focused purposes. Internally also, a significant high proportion of employees in HPOs think that their organization is highly customer-centric or even customer-driven. So, ignoring customer needs can be completely detrimental to sustainable growth of any organization. The most effective vision for change is customer focus. (Ellsworth 2002) He suggests that companies with a customer-focused purpose have:

- found changes easier to manage
- employees who have felt their work more meaningful having got the purpose clear which is customer-orientation
- achieved higher stakeholders returns over long-term
- strong cultures
- more internal alignment

To achieve customer focus, one of the crucial and valuable source for any company is the customer feedback system. Most of the companies which are strategically competitive nowadays have got, if not all, at least some of their operations on-line i.e. through the Internet or WWW. Gathering customer feedback online through e-mails or form-based interfaces is one of the most common activities that companies are engaged in doing on the net, because it gives the customer the flexibility to communicate in an asynchronous domain (which is not the case with the telephone calls) and also gives them a platform to communicate in writing which is a more convenient way as perceived by people for putting the problems or thoughts in a more structured fashion. Also it reduces the anomalies in understanding as everything is in writing, and it also helps the customer or the company to keep a track of the communication that has taken place for example in terms of a series of request-response mails.

This customer feedback information – either in the form of e-mails or some structured textual form-based inputs, is a precious source of knowledge for any organization. In this paper, we are thereby proposing a method for capturing knowledge for change from customer feedback systems using a Knowledge Web.

SOURCES OF CUSTOMER INPUTS/ FEEDBACK

Customer communications with an organization can be broadly classified into two categories based on the purposes of communication: enquiries e.g. on new products/ services/ schemes, or feedback. There are various ways customers communicate to the organizations using technology/ media combinations at different levels of complexities. Figure 1 shows few of such elements.

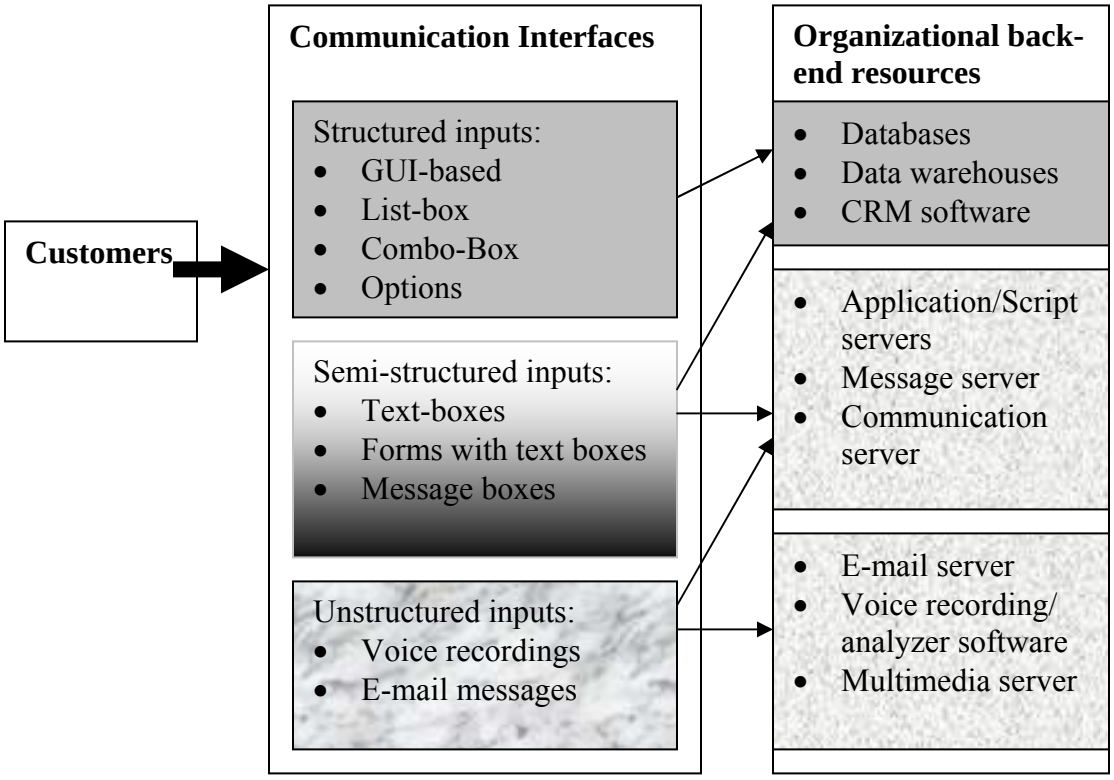

Figure 1: Elements of Customer Communication in Customer-focused organizations

From Figure 1, it can be derived that the customer input/ feedback sources are not only varied in terms of their structured-ness (or unstructured-ness), but also in terms of the underlying technologies like the communication domains, the server technologies, data repositories like warehouses or operational DBMSs. Using all possible combinations of these varieties, capturing the knowledge in a unified form is a daunting task.

EXISTING TOOLS AND TECHNIQUES TO CAPTURE CUSTOMER COMMUNICATION

Generally, in the context of capturing knowledge or discovering reusable patterns from customer inputs, data mining, pattern recognition techniques have been abundantly used. Data warehousing and on-line analytical processing (OLAP) have typically been used to solve data extraction, transformation, data cleaning, storage, and mining issues.

Regarding handling of natural-language based unstructured information sources; previous efforts have used document-based technologies and supported document-level functions such as full text search, document classification, and so on. Business practitioners have developed automated tools to support better understanding and processing of information. Researchers have also developed advanced analysis and visualization techniques to summarize and present vast amount of information. (Fuld et al 2003) Despite recent Improvements in analysis capability (Fuld et al 2003), there is still a long way to go to assist qualitative analysis effectively. Most tools that claim to do analysis simply provide different views of collection of information {e.g.. comparison between different products or companies).Lin (1997) identified various display formats for handling multi-dimensional data e.g. hierarchical displays- an effective information access tool for browsing, network displays, scatter displays (Spence 2001). There are other methods for document representation and visualization also which primarily concerns the task of getting insight into information obtained from one or more documents (Wise et al 1995). Most processes of document visualization involve three stages i.e. document analysis, algorithms, and visualization (Spence 2001). Web content mining treats a web document as a vector of weights of key terms (Bowman et al 1994). He et al. (2001) proposed an unsupervised clustering method that was shown to identify relevant topics effectively. The clustering method employed a graph-partitioning method based on a normalized cut criterion. This method we are using in this paper to extract knowledge from customer communication resources.

CREATION OF A KNOWLEDGE WEB USING CUSTOMER COMMUNICATION ELEMENTS

As explained in Figure 1 regarding various elements in customer communication, there is a need for integrating the knowledge captured from these communication resources. For example, CRM software can capture customer inputs in a very specific structured fashion using list-box options in a form. These inputs can be easily mapped onto a data warehouse or a database. But, customers sending mails or leaving voice messages are also generating valuable knowledge assets. Voice recordings can be converted into text documents. Now, treating these text messages (source being emails/ form based text inputs/ voice messages converted) as unstructured documents, we can use co-occurrence analysis to find the similarities and then consequently the dissimilarities between the messages/ text contents. For this paper, herein after, we are using messages as the unit of customer communication and thereby are referring to any of the communication elements as messages. Messages which are very similar in terms of their contents i.e. many of the identified key-terms (i.e. Terms excluding the general terms like pro-nouns, prepositions, conjunctions etc.) are same, can be clubbed up together to form a cluster. Dissimilar message/ text bodies can be created as other clusters. These clusters can then form a network using hierarchical and partitional clustering method to form a graph which we are mentioning in this paper as a Knowledge Web based on customer communications.

Co-occurrence analysis can convert data indices and weights obtained from messages into a matrix that shows the similarity between every pair of messages say Msg_i and

Msg_j, using similarity measures as follows: (He et al 2002, He and Hui 2002,Shneidermin 1996)

$$Sim_{ij} = \alpha \{A_{ij} / |A|_2\} + \beta S_{ij} / |S|_2 + (1 - \alpha - \beta) C_{ij} / |C|_2 \quad [1]$$

$0 < \alpha, \beta$ (parameters) < 1, $0 <= \alpha + \beta <= 1$, A, S, and C are matrices for A_{ij}, S_{ij}, and C_{ij} respectively. Values for A_{ij} will be 1 if Msg_i has a direct reference to Msg_j, else 0. S is the asymmetric similarity score Msg_i and Msg_j, and is calculated as follows:

$$S_{ij} = sim(Msg_i, Msg_j) = [[\sum_{k=1}^{p} d_{ki} d_{kj}] / [\sum_{k=1}^{n} d^2_{di}]] \quad [2]$$

where n is total number of terms in Msg_i, m is total number of terms in Msg_j , p is total number of terms that appear in both Msg_i, and $Msg_{j.}$, d_{ij} = (Number of occurrence of term j in Msg_i) X $\log((N/d_{fj})Xw_j)$X(Termtype factor); d_{fj} is number of messages containing term j; w_j is number of words in term j; Termtype factor = 1 + ((10-2 X $type_j$ / 10), where $type_j$ = minm 1 if term j appears in subject, 2 if it appears in body, 3 if it appears in 'note' etc.) and C_{ij} is number of messages pointing to both Msg_i and Msg_j.

Once we get the similarity matrices, we create a graph and then partition it to form a network of nodes where the nodes are the representative clusters of a group of messages having high similarity scores among them. Partitioning of a graph, say G, can be done in various ways, for example, by using Normalized cuts: (Rich and Knight 2001,Shi and Malik 2000).

Normalized Cut (x) = {cut between (A, B)/ assoc(A, V)} + {cut between (A, B)/ assoc (B,V)} [3]

where, Cut between (A,B) = $\sum_{i \in A, j \in B} Sim_{ij}$, Sim_{ij} is similarity between nodes i and j of the graph. Assoc(A,V) and assoc(B,V) shows how on average nodes within a group are connected to each other. A cut on a graph G = (V, E) is defined as removal of a set of edges such that the graph is split into disconnected sub-graphs.(Chen et al 1998,Chen et al 2001).

The process is explained the following Figure 2.

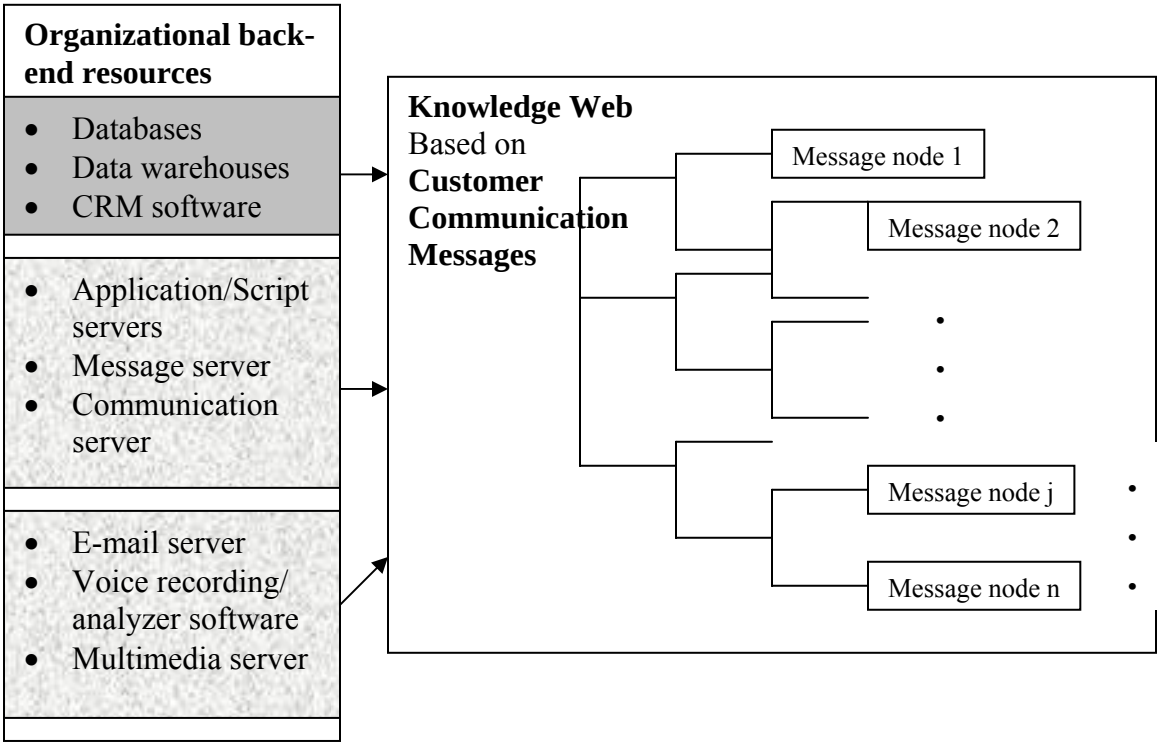

Figure 2: Creation of Knowledge Web using Similarity Scores and Graph Partitioning

The steps in Figure 2 are explained as below:

1. Depending on the message type i.e. structured/ semi-structured or unstructured, a textual representation will be created as follows:

    a. If the message is stored in a database as structured inputs from a GUI-based options-selection form, then

        i. data-values in each row of the database has to undergo a de-codification, i.e. if the data values are indexed/ annotated/ coded, then the meaning of their annotations and coding will be replacing the index value.
        ii. These de-codified field-values will be concatenated to form a string.
        iii. This string will be treated as an input message unit to the knowledge web module.

    b. If the message is semi-structured type, the structured and unstructured field in the message will be combined/ concatenated to form a text string to be treated as an input message for the knowledge web module.

c. If the message is unstructured i.e. an email message or text-box entry, the message will combine the subject, notes and other fields together to represent a unified message input format.

2. Now, these messages in a unified text format will be input to the Knowledge web module

3. Say, Msg_i arrives for feed into the module.

4. Similarity scoring module measures the similarity score of the message Msg_i (for $i = 1$, $S_{ij} = 0$) with respect to all existing nodes say $MsgNode_{i=1 to n}$

   a. If the similarily score for Msg_i >= a given similarity threshold value for data reduction for any existing MsgNode say $MsgNode_j$, then the KM_j gets mapped or included into the node $MsgNode_j$

   b. if similarity score is < threshold value, a new node for Msg_i is created.

The steps are explained with an example message Msg_i in Figure 3 given below:

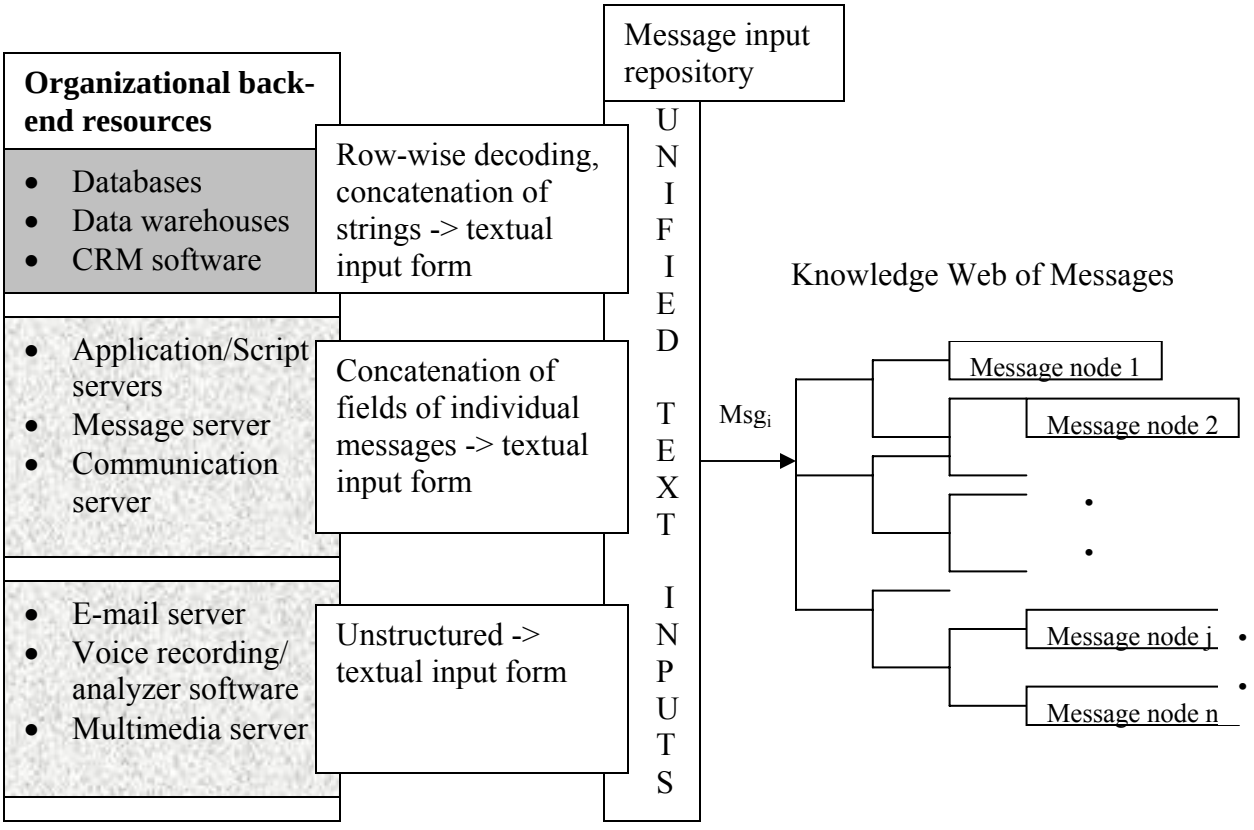

Figure 3: Steps in Creation of Knowledge Web using Similarity Scores and Graph Partitioning (An Example)

If the Maximum similarity score between Msg_i and $MsgNodes_{i=1to\ n}$ >=Threshold, then Msg_i gets included as shown in Figure 4, in the $MsgNode_j$ which has that maximum similarity score with Msg_i

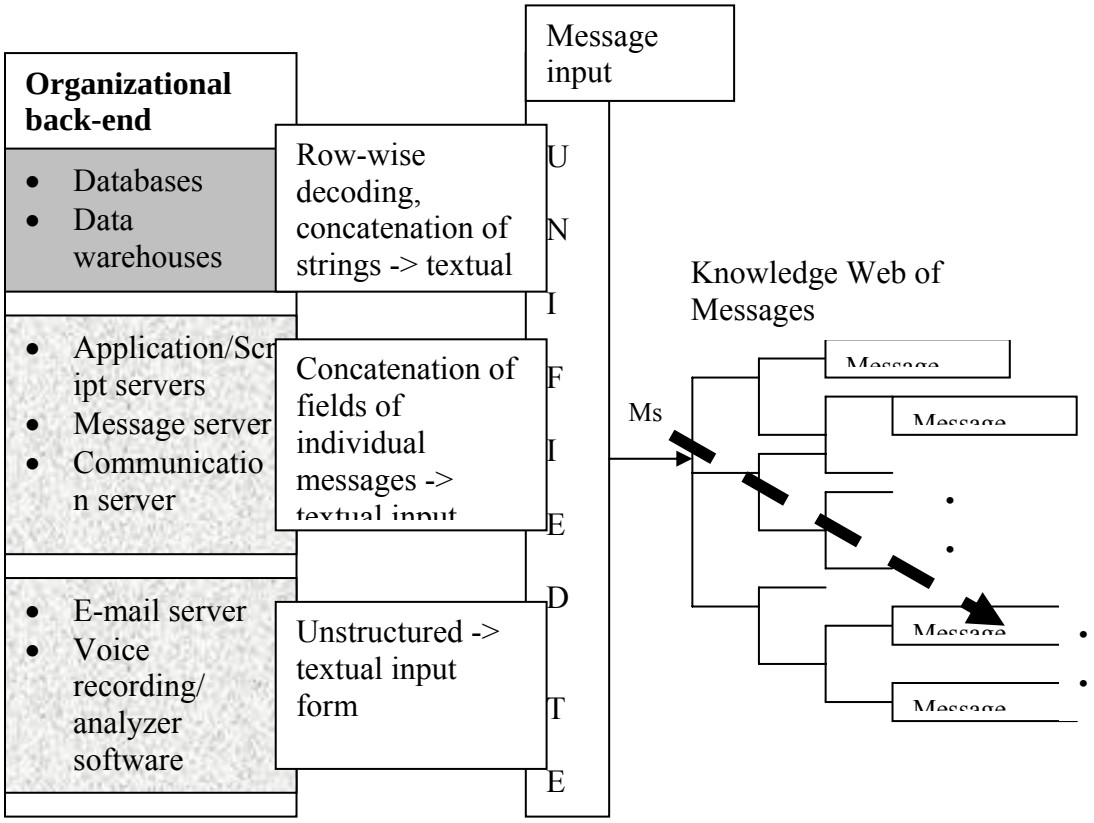

Figure 4: Absorbtion of Message into Message Node

The Msg_i gets absorbed/ included in $MsgNode_j$ which is having maximum similarity score, and Maximum similarity score between Msg_i and $MsgNodes_{i=1to\ n}$ >=Threshold.

Alternately, if Maximum similarity score between Msg_i and $MsgNodes_{i=1to\ n}$ <Threshold, then Msg_i creates a new node in the Knowledge Web because it has minimum similarity wth the existing nodes and therefore can contribute as a new node. So, in this case the Knowledge web gets enhanced as shown in Figure 5 below:

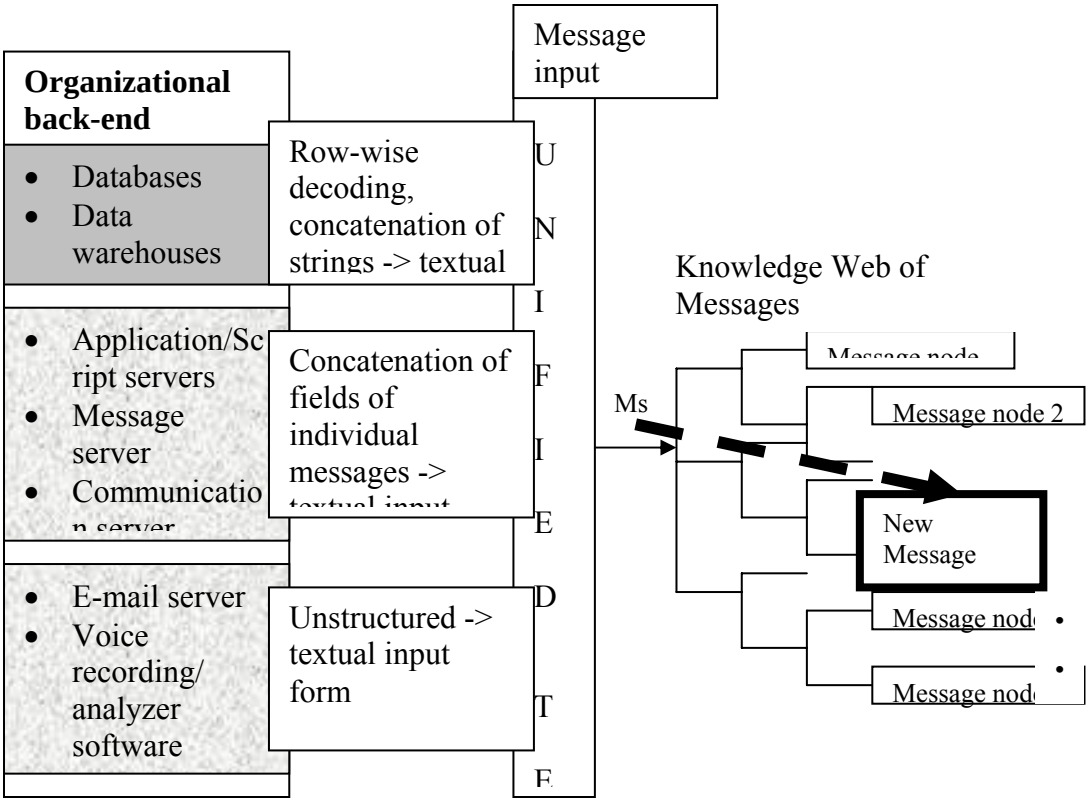

Figure 5: Creation of a New Message Node with Msg_i

This process has two primary advantages:

1. It integrates all customer communication elements in a unified domain represented my messages
2. This collective repository of customer communication can now be interfaced with various legacy analytical systems, OLAP systems, mining software tools etc. for providing support in decision making in multiple domains or specific purposes like customer profiling, product/ promotion design, designing advertisements, customer segmentation, understanding customer requirements, discovering hidden patterns in customer behaviors by getting the high-dimensional network analyzed with mining tools etc.

CONCLUSION

The knowledge web created with inputs from a unified repository of various elements of customer communication, as explained in this paper, can be the most valuable resource an HPO can develop and nurture, to gain competitive advantage and further insights into customers' minds, and the market at a macro-level. Use of this web can thereby need not be restricted only to an IT-enabled, strongly technical, analytical domain but it can be expanded to even strategic decision support with suitable visualizing and reporting tools developed and integrated with this precious knowledge repository. The idea can be

expanded, tried and tested with other knowledge representation models also, for example suing the UIMA(Unstructured Information Management Architecture) proposed by IBM. However, the basic requirements of such a system, i.e. integration of various communication elements, and preserving their semantic value, have to be supported adequately by the other possible alternatives.

REFERENCES

Bharat, K., Henzinger. M.R., *Improved algorithms for topic distillation in hyperiinked environments*. Proceedings of the Twenty-First International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press. 1998. pp. 104-111.

Blum. C., Roli. *A. Metaheuristics in combinatorial optimization: Overview and conceptual comparison*. ACM Computing Surveys. 35, 3 (2003). pp.268-308.

Bowman. C.M, Danzig. P.B., Manber. U.; Schwartz, F'. Scalable Internet resource discovery: Research problems and approaches. *Communication of the ACM. Vol* 8 (1994). pp 98-107.

Chen. H.; Chung, Y.; Ramsey. M.; and Yang. C. A smart itsy bitsy spider for the Web. *Journal of the American Society far Information Science. 49.* 7 (1998), 604-618.

Chen, H.; Fan. H.; Chau. M.; and Zeng, D. Meta Spider: Meta searching and categorization on the Web. *Journal of the American Society for Information Science and Technology. 52,* 13(2001), 1134-1147.

Collins J., *Good to Great*, Harper Collions Publishers Inc,NY,2001.

Dixon N. M., The Organizational Learning Cycle: How We Can Learn Collectively. New York: McGraw-Hill, 1994.

Duncan R., Weiss A., "Organizational learning: Implications for organizational design," in Research in Organizational Behavior, B. Staw, Ed., vol. 1 Greenwich, CT: JAI, 1979, pp. 75–123.

Ellsworth,R.R, *Leading with Purpose: The New Corporate Realities,*Stanford University Press,Stanford,2002.

Fuld, L.M.: Singh. A.: Rothwell. K.; and Kim, J. *Intelligence Software Report™ 2003: Leveraging the Web.* Cambridge. MA: Fuld & Company, 2003.

Global Solutions Directory, Text Mining Server, http://www.developer.ibm.com/solutions/isv/igssg.nsf/list/bycompanyname/86256B7B0003EBBF86256DF600491B20?OpenDocument.

Futures-Group Ostriches & Eagles. The Futures Group Articles, Washington, DC, 1997 (available at www.futuresgroup.com).

Global Solutions Directory, Text Mining Server, http://www.developer.ibm.com/solutions/isv/igssg.nsf/list/bycompanyname/86256B7B0003EBBF86256DF600491B20?OpenDocument.

Hackbarth G., Grover V., "The knowledge repository: Organizational memory information systems," Information Systems Management, vol. 16, no. 3, pp. 21–30, Summer 1999.

Halbwachs M., The Collective Memory (F. J. Ditter and V. Y. Ditter, Eds. Translation of 1950 Original). New York: Harper Colophon, 1950/1980.

He. X.; Ding. C; Zha. H.; and Simon, H. Automatic topic identification using Webpage clustering. In X. Wu. N. Cercone, TY. Lin, J- Gehrke. C. Clifton. R. Kotagiri. N. Zhong. and X. Hu (eds,). *Proceedings of the 2001 IEEE International Conference on Data Mining.* Los Alamitos. CA: IEEE Computer Society Press. 2(X)I. pp, 195-202.

He, Y, and Hui. S.C. Mining a Web citation database for author co-citation analysis. *Information Processing and Management. 38.* 4 (2002). 491-508.

Holbeche L.,*The High Performance Organization*,Elsevier Butterworth-Heinemann,2005.

Hyung Jun Ahn; Hong Joo Lee; Kyehyun Cho; Sung Joo Park. Utilizing knowledge context in virtual collaborative work. By: Decision Support Systems, Jun2005, Vol. 39 Issue 4, p563, 20p,

Iannella R., "An Idiot's Guide to the Resource Description Framework," *The New Review of Information Networking* **4 (**1998), http://www.dstc.edu.au/cgi-bin/redirect/rd.cgi?http:// archive.dstc.edu.au/RDU/reports/RDF-Idiot/.

Kealy, W.A. Knowledge maps and their use in computer-based collaborative learning. *Journal of Educational Computing Research, 25.* 4 (20()l). 325-349.

Krippendorff K., "Some principles of information storage and retrieval in society," General Systems, vol. 20, pp. 15–35, 1975.

Lin, X. Map displays for information retrieval. *Journal of the American Society for Information Science. 4H.* 1 (1997), 40-54.

Lyman. P., and Varian. H, How much information. University of California, Berkeley, 2(KX) (available at www..sims.berkeiey.edu/how-much-info).

Mowshowitz. A., and Kawaguchi. A. *By Online Web. Communications of the ACM, 45.* 9(2002), 56-60.

Nevoa, Dorit; Wand, Yair. Organizational memory information systems: a transactive memory approach. Decision Support Systems, Jun2005, Vol. 39 Issue 4, p549, 14p,

Nucleus Report on Top 10 IT Spending for 2005: Survey of CIOs in MNCs: Survey Report March 2005 by Nucleus Research, http://www.nucleus.com/surveys/2005

Rich E., Knight K., *Artificial Intelligence*, Tata McGrawHill Publishing Company Ltd, N. Delhi, 2001.

Rowden,R.W., "The Learning Organization and Strategic Change",SAM Advanced Management Jl,Vol.66,No.3,2001.

Shi. J., and Malik. J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence. 22.* S (2(X)0), 8S8-905.

Shneidermiin, B. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of IEEE Symposium on Visual languages.* Los Alamitos, CA: IEEE Computer Society Press, 1996. pp. 336-343.

Spence, R. *Information Visualization.* New York: ACM Press, 2001.

Torgerson, W.S. Multidimensional .scaling: 1. Theory and method. *P.sychometrika. 17.4* (1952), 401,19.

Vapnik V. N., *The Nature of Statistical Learning Theory*, Springer-Verlag New York, Inc, New York (1995).

Wise, J.A.; Thoma, J.J.: Pennock. K.; Lantrip, *D.:* Pottier, M.; Schur, A.; and Crow, V. Visualizing the non-visual: Spatial analysis and interaction with information from

text documents. In *Proceedings of IEEE Symposium on Infonnation Visualization.* Los Alamitos, CA: IEEE Computer Society Press, 1995, pp. 51-58.

Young, F.W. *Multidimensional Scaling: History, Theory, and Applications,* ed. R.M. Hamer. Hillsdale, NJ: Lawrence Erlbaum, 1987.

Zanasi A., "WebMining through the Online Analyst," in *Data Mining II*, N. F. F. Ebecken and C. Brebbia, Editors, WIT press.com electronic library (2000), pp. 3–

Z. Su, L. Zhang, and Y. Pan, "Document Clustering Based on Vector Quantization and Growing-Cell Structure," *Proceedings of the IEA/AIE 2003, Developments in Applied Artificial Intelligence, 16th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, Laughborough, UK; Lecture Notes in Computer Science 2718 Springer (2003), pp. 326–336.