Information Overload in organizations
A New Model to Diagnose

Mónica del C. Blasco-López[1, 2*], Robert Hausler[1], Rabindranarth Romero-López[2], Mathias Glaus[1]

[1] Station Expérimentale des Procédés Pilotes en Environnement, École de Technologie Supérieure, Université du Québec, Montréal, Canada
[2] Unidad de Investigación Especializada en Microinformática y Tecnología Ambiental, Facultad de Ingeniería Civil, Universidad Veracruzana, Xalapa, Veracruz, México
monica-del-carmen.blasco-lopez.1@ens.etsmtl.ca, robert.hausler@etsmtl.ca
rabromero@uv.mx, mathias.glaus@etsmtl.ca

Abstract

People inside organizations have a daily fight with big quantities of information which becomes to be the problem so called Information Overload. The solutions for this problem have been to acquire new technologically advanced devices to deal with the problem, machines which can process larger quantities of documents with the same information inside. The concern of this research is the analysis of the efficiency of the document containing the information, using a new model Analogic – Conceptual - Thermodynamic (ACT) to diagnose if data has information overload and once detected, be able to contribute to reduce the information load in the system. This paper outlines an approach taking an administrative format as an example. As initial finding it has been development ACT which is an arrangement for data classification based on the calculation of informational entropy and enthalpy. This new model is a new simple and practical tool that can be used by managers in organizations to assess their formats in use to verify and to diagnose if there is, or there is not, an information overload.

Introduction

In this information age which we live, handle information is not an easy task. In a recent survey of 100 government workers (from federal, state, and local organizations), 70% indicated that they "sometimes" or "always" feel overloaded by information [1].

The information overload problem affects individuals, organizations and society by the same, this phenomenon was defined at the beginning by some researchers simply as too much information [2] others affirm that this is not only caused by having too much information but it is a complex phenomenon resulting from others factors such as uncertainty, ambiguity, novelty, complexity or intensity [3]. [4] indicate for their part, that the information overload occurs when the received information begins to block more than to help.

The information has a relation in whole, with the data it contents, people who drive it, objectives or purposes for which it is used in the organization, the source that generates, the canal used for transmission, its quantity and its quality. [5] point out that the principals components of information overload are: volume, time constraints, noise and channels, the latter being little analyzed so far.

There are very few proposals addressing the information overload problem that contemplate their cause and not the effect, and even fewer models that provide practical solutions to this problem. [6] for example, to analyze the problem by considering the factors that [2] cause the effects of too much information on individuals. Most of the proposals are related to the person who suffering the effects and not according to the object that contains it. Therefore, the exploration on this unknown topic was performed in this research.

Discussion and hypothesis

The reflection leading this analysis is why not instead of acquiring big readers or scanners that process at high speeds thousands of data in organizations, it chooses to analyze and verify forms containing these data, such objective will be to reduce the amount of these from the beginning, before being processed and even without going to expensive equipment that only mess more noise in the system?

This research part from the assumption that the information overload problem arises in the quantity and quality of the data generated from the source, phenomenon sometimes given for the lack of clarity of the information requested on the form and sometimes for the processor need to have all the necessary (and unnecessary) information on hand to do their job, causing a drop in the performance or efficiency of the form which ultimately leads to an extra expenditure of energy, for example in human or materials resources and this accumulation of extra-energy can be translated as information overload.

This model is based on the hypothesis of considering an administrative form as a thermodynamic system and its objective is to develop a model to diagnose if in the forms used in organizations are data which produce information overload in the system.

Whit this objective, the remainder of this document has been organized as follows: the theoretical principles of the model are presented. Then the analogy is explained, this is followed by the description of the model and explained by an example and finally, the results and a series of conclusions with managerial implications are presented.

Theoretical principles of the model

The Analogic-Conceptual-Thermodynamic (ACT) model was named this way because of from the analogy among the information system flow in an organization and a thermodynamic system in order to detect if there is information overload in the administrative forms that flow daily between different departments.

The thermodynamic theory studies the state and general behavior macroscopic of a system [7]. This approach has been used in other branches of knowledge (to study for example the behavior of different systems of the human body, the urban transport, the economic system, dams, production systems) [8] [9] [10] to provide solutions to daily problems.

The first two principles of thermodynamic have been taken as a basis for the development of this proposal. The first principle deals with the conservation of energy and the second one deals with its evolution introducing this way the entropy.

The subject of the entropy, in the communication system, in general, refers to the mathematic model of the communication developed by [11], who related the communication system with the probability of choose a letter, a symbol or message in the communication. Shannon states that communication can be treated as an engineering problem from a syntactic point of view, which is to say from the symbols that conform it, the language has a structure pre-stablished and a finite number of elements or symbols that conform it.

From these theoretical principles, an analogy between a thermodynamic system and information flow in administrative system has been made, taken a paper form as channel of communication and container of data that flow.

The Analogic-Conceptual-Thermodynamic (ACT) Model

Briefly, the analogies made will be described to better understand the operation of the model. As starting point, the conservation of energy principle may be expressed as follows:

The net change (increase or decrease) in the total energy of the system during a process is equal to the difference between the total energy entering and the total energy leaving the system during that process [12].

$$E_{IN} - E_{OUT} = \Delta E_{system} \qquad \text{Eq. 1}$$

The model is based on a simple balance observation, for example in the human being, energy enters as food, and later this energy is transformed into work or energy storage. In the same way, in an administrative form energy enters represented by data that the applicant fills. Later, this energy is transformed in information or data storage. The difference between "data" and "information" is made by the signification that some data have for the receptor. If the data do not mean anything for the receptor, these remain in a state as simply "data", but if these data mean something for the receptor, it generates one reaction in the receptor, which is to say that it generates a work, then these data are transformed into "information" [13].

Figure 1. The analogy between the energy balance in the human being and the energy balance in administrative format.

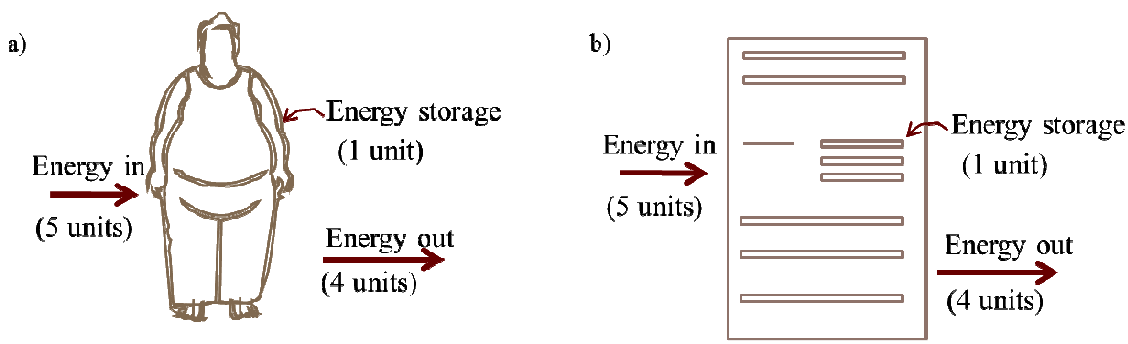


Figure 1. a) Energy balance in the human being. Figure adapted from (Çengel & Boles, 2002) b) Energy balance in administrative format.

Then, the internal energy in the system is represented by the following equation:

$$U = W + Q \qquad \text{Eq.2}$$

Where U is the internal energy, W represents work and Q represents heat transfer. Analogically, following the first principle of thermodynamic, a machine can transform heat transfer into work although not all of this heat is transformed into work, the heat unconverted stays as heat storage.

Thermodynamically, this useful work is known as enthalpy, so hereafter, in this research, it will be called Informational Enthalpy and it will be represented by H_I, and the remaining data will be represented by Q_I. Resulting the formula of the internal informational energy as following:

$$U_I = H_I + Q_I \qquad \text{Eq. 3}$$

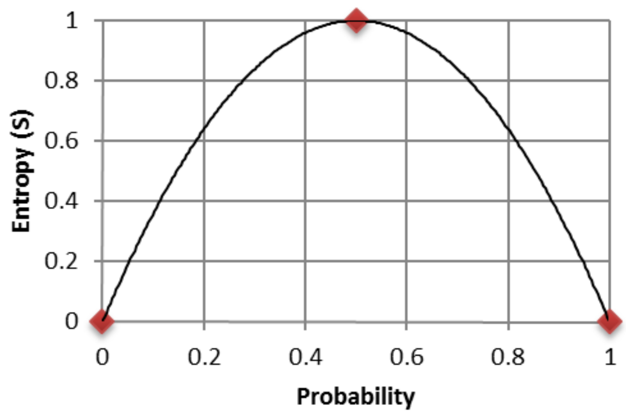
The first law of the thermodynamic said that exist a certain amount of energy in a system and that it can be transformed. The second law states that there is not system 100% efficiency and that it is because there is something called entropy. This property deals with the disorder in the system. Shannon (1948) observed this similitude in the information system and he states his equation of informational entropy, which is an average amount of missing information in the system.

Then, this proposal considers the informational entropy too, and it is calculated from the Shannon equation:

$$S = -K \sum_{i=1}^{n} pi \log pi \qquad \text{Eq.4}$$

Where *K* is a positive constant, *pi* is the probability of the occurrence of *I* and *S* is the entropy. Shannon originally uses the letter *H* to designate the entropy of the information but for this analysis, the letter *S* will be used to avoid future confusion with the enthalpy (*H*) defined above. The statistical entropy (S) points out that if the system is perfectly known, or otherwise, perfectly unknown, the information is at maximum and, therefore, the value of the entropy is equal to zero. On the other hand, if there is an uncertainty about the behavior of the system, the entropy takes a value different than zero ($S \neq 0$). Then it is stated that entropy (S) measures the missing information of a system: the bigger is S, the more there is uncertainty in the system considered. The graphic 1 shows the behavior described above.

Graphic 1. The behavior of the informational entropy (S) in the system.



Graphic 1. Adapted from (Shannon, 1948)

Recapitulating, the first law talks about the energy conservation and the second law talks about his evolution affected by the entropy. For all that the ACT model proposes the follow equation for detecting the data that generate the overload in the information.

$$U_I = H_I + Q_I S \qquad \text{Eq. 5}$$

That is, the informational internal energy is equal to the informational enthalpy plus the remaining data affected by its entropy.

The following, the steps of the ACT model will be described briefly.

Steps of the Analogic-Conceptual-Thermodynamic (ACT) model

For clarifying the model, a fictional example was taken. This consists in one format that flows for three exchange stations or different departments in an organization.

1. Format structure

Figure 3. Field = case

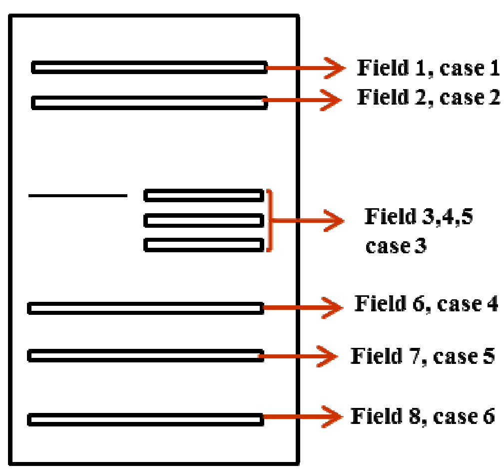


Figure 3. From the analysis of the general structure of format, the number of cases of the system is extracted.

In first, the format structure is analyzed; an identification number is given to each field. The administrative format is seeing as the whole analyzed system. For example, a format with 15 fields to answer; will have 15 different cases. No matter what kind of answer has each field because this model does not analyze if the answer is black or white but it analyzes the useful of this data for the whole system.

2. Calculation of probability and informational entropy

To work with the informational entropy, first is needed to calculate the probability of each case. The ACT model works from a binary code, where *I* represents useful data for information processing in any of the exchange station while the format transit and *O* represents not-useful data in any of the exchange station. In table 1 are the data for the example that help to clarify this step:

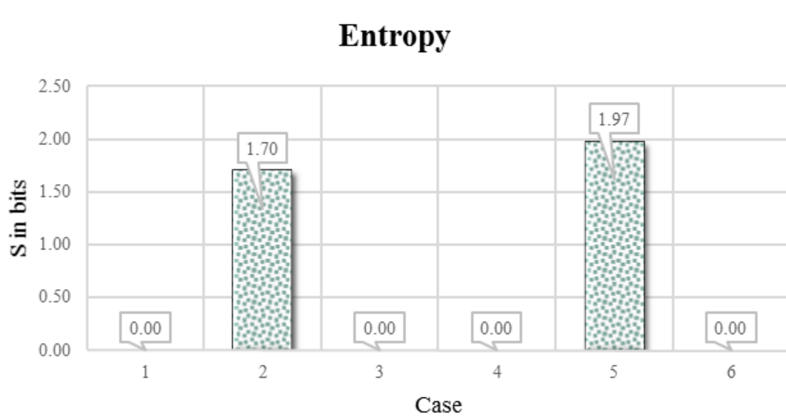
Table 1. Probabilities for each case.

Case C	**Probability I P(I)**	**Probability O P(O)**
1	1.0	0.0
2	0.7	0.3
3	0.0	1.0
4	0.0	1.0
5	0.3	0.7
6	1.0	0.0

Table 1. It presents probabilities of useful and probabilities of useless of each case.

With the probability calculated, from equation (3) is possible to calculate the informational entropy. For this example, the results are presented in graphic 1. It is easy to see that the case with probability I [P(I)] equal to zero or one, have an informational entropy equal to zero and the other data with 0 < P(I) > 1 the entropy has a variable value.

Graphic 2. Value of informational entropy of each case.



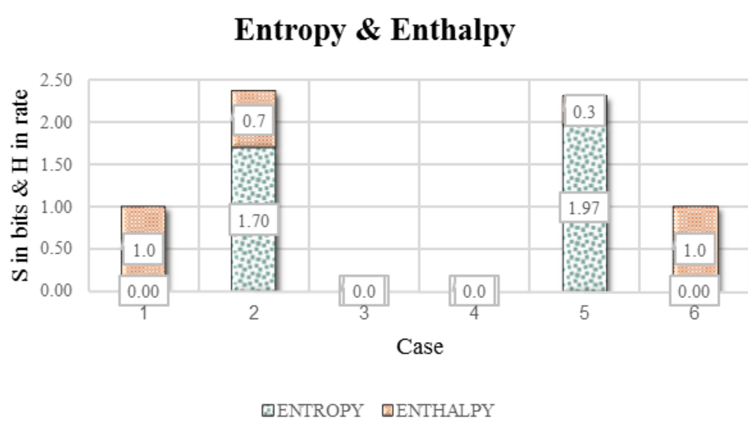
The informational entropy is not the only property to be considered for determining if a data “x” is a potential generator of information overload, also the calculation of informational enthalpy is needed to be able to assess the situation.

3. Calculation of the informational enthalpy.

As it was mentioned previously, in the ACT model, the informational enthalpy corresponds by analogy to the energy that produces useful work in the system. Then, for this model the informational enthalpy makes reference to the data that produce true information. In this calculation, the informational enthalpy (H_I) corresponds to P(I).

So, if these two values are plotted on the same graph, it is clearly visible that data with entropy equal to zero (because there is a certainty about them), there are some of them totally useful and other ones totally useless. Being these last the generators of the problem of information overload from the root.

Graphic 3. Informational entropy and informational enthalpy in the same graph.



4. Calculation of the informational internal energy

Already having the calculation of the probability, the entropy and the enthalpy of the system, now only there remains to fulfil the last step of the ACT model, the calculation of the internal informational energy of the system (U_I) from the equation 4.

The informational enthalpy represents the data which make the information; it says that it is not uncertainty of this part of the system. Now in the system stay the remained data, which represents some kind of energy with some grade of uncertainty, so each case of this remained data is multiplied by its own degree of missing information, it is to say, its entropy. All this is made in order to give the informational internal energy of each case and thus, proceed to its final evaluation.

Table 2. Calculation of informational internal energy of each case

Case	H_I	Q_I*S	U_I
1	1.00	0.00	1.00
2	0.70	1.70	2.36
3	0.00	0.00	0.00
4	0.00	0.00	0.00
5	0.30	1.97	2.31
6	1.00	0.00	1.00

Results

What can be seen in the above table is that there are data with U_I equal to zero, one or major to zero but different to one. What these values mean according to the analogic interpretation?

For easier interpretation, the data have been classified into three different zones according to its value. Following, the description of each zone will be made:

Data with a value of $U_I = 1$ is the result of the addition of enthalpy to one plus entropy equal to zero. This means these data are indispensable for the format processing. There is a full certainty to use this data, is to say, these data represent 100% of the information and therefore 0% of uncertainty. For that reason, this zone is called the *zone of load of information*.

The second type of data, these data have an internal informational energy equal to zero. This is the result to sum an informational enthalpy equal to zero plus the informational entropy equal to zero too. These data represent the data that is unnecessary in the format and at the end; they generate the excess of information in the system. It is possible to process the format without of these data. When removed, the system continues working and the system becomes more efficient. The zone where such data fall is called: *zone of information overload*.

And finally, there are data with $U_I \neq 1$or $U_I \neq 0$. These data fall in the zone called: *zone of remaining data* because these data have an uncertainty about their usefulness. These data have a variable entropy and enthalpy.

Table 3. Summary of classification zones according the ACT model

Zone	U_I
Load of information	$U_I = 1$
Information overload	$U_I = 0$
Remaining data	$U_I \neq 1$or $U_I \neq 0$

From the case taken as example, 33% of data fall in each zone. It means that a third (this value change depends on each system) of data is the potential generator of the information overload problem with which the most of the people inside organizations have to deal daily.

Conclusions

The purpose of this paper was to demonstrate that many times when a global problem is analyzed, much emphasis is made on the effects of the problem but only little bit attention is given to the local causes.

The study realized by (Farhoomand A. F., 2002) proposes in the first place as solutions to deal with the problem of information overload in organizations: filter the information. The aim of this research is precisely to offer a practical tool to filter the enormous quantity of data and information which organizational workers have to deal every day.

The ACT model here proposed provides not only the fact to detect potential data generators of information overload but also it is useful to detect the minimum amount of data necessary to process one format.

With this tool, department managements will be able to assess the format that flows daily in his/her department and to diagnose if it is needed a re-engineering of them.

Implications

This new model can be considered as one point of depart for future explorations made from analogies between thermodynamic systems and administrative information flow systems.

In this time, it is possible to observe the big difference that can be made when 33% of the potential energy that data have inside them.

Imagine that this 33% of example data be real. What or how much of these data represent for the organization? In an isolated manner, probably it is no big deal but if these two data are multiplied by 100, 1000 or 10,000 formats that flow daily among departments, then the problem becomes into one of big dimensions.

Then it occurs something similar to the human being, the accumulation of energy becomes to be visible and the local problem turns into a global problem.

References

1. Clarke, R. Y., & O'Brien, A. (2012). The Cost of Too Much Information: Government Workers Lose Productivity due to information overload. Virginia, United States: IDC Government Insights.
2. Eppler, M., & Mengis, J. (2004). The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society: An International Journal, 20:5*, 325-344.
3. Schneider, S. C. (1987). Information Overload: Causes and consequences. *Human Systems Managemnt, 7*, 143-153.
4. Bawden, D., Holtham, C., & Courtney, N. (1999). Perspectives on information overload. *ASLIB Proceedings, 51*(8), 249-255.
5. Farhoomand, A. F., & Drury, D. H. (2002). Managerial Information Overload. *Communications of the ACM, 45*(10), 127-131.
6. Jackson, T. W., & Farzaneh, P. (2012). Theory-based model of factors affecting information overload. *International Journal of Information Management, 32*, 523-532.
7. Sanz, M.-N., Badel, A.-E., & Clausset, F. (2003). *Physique Tout-en-un 1er année (2e ed).* Paris: Dunod.
8. Rojo, G., Glaus, M., Hausler, R., Laforest, V., & Bourgois, J. (2013). Dynamic Waste Management (DWM): Towards an evolutionary decision-making approach. *Waste Management & Research., 31*(12), 1285-1292.
9. Valero, A., Valero, A., & Calvo, G. (2014). Using thermodynamics to improve the resource efficiency indicator. *Resources, Conservation and Recycling, 94*, 110-117.
10. Vega-Azamar, R. E., Romero-López, R., Glaus, M., Oropeza-García, N. A., & Hausler, R. (2015). Sustainability Assessment of the Residential Land Use in Seven Boroughs of the Island of Montreal, Canada. *Sustainability*, 2454-2472.
11. Shannon, C. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 379-423.

12. Çengel, Y. A., & Boles, M. A. (2002). *Thermodynamics, Fourth Edition.* New York: McGraw-Hill Higher Education.

13. Alain, G. (2007). L'information non structurée dans l'entreprise: usages et outils. Paris: Hermes science: Lavoisier.

Acknowledgment

Authors tanks to École de Technologie Supérieure (ÉTS), Universidad Veracruzana (UV), Professorship Development Program (PRODEP) and Secretary of Education of Veracruz (SEV) for its support and fully appreciates any suggestions or notes regarding this work, at the following correspondence address: monica-del-carmen.blasco-lopez.1@ens.etsmtl.ca