Information Dynamics: Understanding Information in Computer and Software Systems*

Christian B. Almazan
University of Maryland, College Park
almazan@cs.umd.edu

Abstract

We present an information-oriented approach to system design and maintenance called Information Dynamics. This differs from current process-oriented approaches by understanding the vital role information has in any system, whether it be for business, government, academic, etc. Through looking at information primarily, Information Dynamics will give all of the people involved in system development a fresh look at the important aspects of the system.

1 Introduction

The development and maintenance of today's computer and software systems compromises the majority of information technology costs, time and money. We believe that the methods used in creating such systems may be one reason why it takes so many resources for these systems. Traditional and modern software engineering processes often take a process-oriented approach to solving problems. While most business operate with a well-defined process model, the creation of software systems to mimic them should be different. We present an information-oriented approach called Information Dynamics which augments the traditional process-oriented approaches.

In the Information Dynamics approach, we take information to be the primary component of any system, not just internal to the processes. By making information first class objects in a system, we allow for easy extensibility and understanding of how a system should be built and maintained. In order for an information-oriented approach to work, the following questions must be considered:

- Who needs what information?
- When does someone need information?
- Where does the information exist?
- How do we process this information?
- Does this information satisfy applicable needs?
- Does a process need more or less information?

*Submitted in order to fulfill the requirements of the Master of Science in Computer Science without Thesis option for the Spring 2005 semester. Professor Ashok K. Agrawala advised the author.

These questions must be answered since information comes in at a tremendously high rate and the proper information must be processed. The rest of this paper outlines the key ideas of Information Dynamics. It should be taken as a guideline in order to understanding how information plays a vital role in computer and software systems. The invention of the computer would not have happened if we did not have the need to process large amounts of information.

2 Defining Information

The term information cannot be considered an easy word to interpret. With several definitions of the word itself and multiple related entries in the dictionary[1], how should information be defined while being used in the Information Dynamics framework? We present several different interpretations of the term information briefly in this section and will discuss what we consider useful information in our context.

In the first chapter of his dissertation work, Timpson [18] goes to great lengths in distinguishing two types of information: everyday information and technical information. Everyday information can be seen as how we communicate on a daily basis, whether it be verbally, through e-mail, writing letters, reading a book, etc. This type of information has semantics and meaning attached to it. He differentiates this type of information from technical information since the remainder of his thesis work primarily concerns technical information in the quantum information sense. Timpson relates technical information to Shannon information [17]. Shannon information, which gave rise to information theory, describes the amount of information contained within communicating parties. This information includes the number of bits transferred and what kinds of compression and error correction can be performed on the messages. Technical information does not concern itself with the semantics and meanings of the messages as it has no relevance to the task at hand, the quantity of information attainable. Looking at these two definitions, Information Dynamics can be seen as looking at everyday information, as we will see in later chapters. We should avoid removing information theory from our arsenal, since it can be useful in understanding how communication works. However, autonomous entities do not care about how communication mechanisms works, but whether or not the appropriate information has been given to them at the right time or at the right location.

In addition to Timpson, numerous articles have been written trying to define information. We take three of these articles and describe their main points. These articles only agree upon one thing, the title: "‘What is Information?"’ We will consider articles from McGovern, Chmielecki, and Israel and Perry.

McGovern [15] first describes the differences between information and resources. Unlike resources, information cannot be consumed, but only shared among others and information can only be reproduced with an enormous cost. He then goes on to attempt to find definitions of information in several dictionaries and comes up with a major overall theme, that the complete opposite of information is chaos. He comes to this since information has form and structure, process and action, and communicates knowledge and intelligence. In our framework, we consider the movement of information paramount to the success of a system. If McGovern's propositions are true, then information movement must be optimized because of the expensive nature of the reproduction of information.

[1] http://www.m-w.com/cgi-bin/dictionary?book=Dictionary&va=information

Chmielecki [4] considered a different definition of information. His definition can be likened to that of a turing machine or biological codes of DNA and RNA since he considers information to be contained within collections, alphabets, codes, repertoires, and information. An important distinction he indicates in his work describes distinguishing information contained within different evolutionary stages of an organism: parainformation, structural information, and metainformation. We indicate a similar distinction in section 3 which distinguishes information from its representation.

Israel and Perry [10] [11] indicate that information has several different parts when dissecting natural language. By doing this, it allows multiple pieces of information to be analyzed separately or fused together. In particular, they describe several different properties of information. An important property, situations, must be paid particular attention to. Information can only be considered useful if used in the context of a particular reality. We will see that contextual information, in the form of models, plays an important role in Information Dynamics.

One final consideration in understanding information comes from Floridi's work on the philosophy of information [7]. This work, considerably different than the philosophy of science work done by Timpson, considers information in the context of the technological world. With the advancement of computers, whether it be the availability of computing devices, artificial intelligence, etc., information now has to be understood at a very conceptual level, which includes understanding the basic principles of information and what application does information have to philosophical problems. He describe several different concepts, some of which have not been described in this paper, but an important one deals with the dynamics of information. Three key items have been defined: the constitution and modeling of information environments, information life cycles, and computation. We will see these key items have a place in Information Dynamics and it will be interesting to see, in the future, how the span of Floridi's work relates to Information Dynamics.

3 Representing Information

Understanding information in Information Dynamics requires more than just understanding what information provides to us, but what limitations information imposes upon a system. We believe that information has at least two forms: the informational level and the representational level. Entities, including sentient beings such as humans, manipulate information in their brains. Such information does not have a direct representation outside of the brain (or scientists have not been able to pinpoint the exact structure of information in the brain yet). In order to communicate with other entities or impose information outside of their own brains, entities have to translate information into an appropriate representational form.

An appropriate example to show how representation takes part in our everyday lives can be seen easily through basic numbers and counting. When growing up, children often learn the numeric alphabet, the numbers zero through nine. Depending on where the children grow up, they could learn these numbers in any number of ways, including Arabic numerals, 0 to 9, or a Roman numeral style, I to IX. On the other hand, computers do not understand these at a basic level, they only understand binary, 0000 to 1001. The list continues, from Morse Code to the abacus.

Even with all of these different numerical alphabets, entities can easily map one representation to another. Commonly, school children will be taught how to convert the Arabic numerals to Roman numerals and vice versa. Such representations have a one-to-one correspondence in which nothing will be lost. However, when translating from the informational realm to the physical representations, there may be a loss of information. This type of information does not mean the

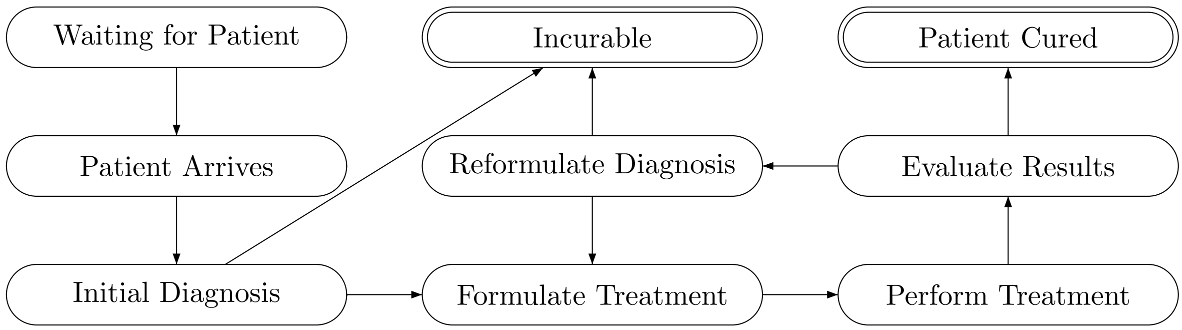

Figure 1: Doctor's Process View of Treating a Patient

same as technical information, we lose semantic information and not the amount of information. For instance, when a translator translates from one language to another, there may be some loss due to the fact no two languages will be the same. Furthermore, different translators may have different interpretations of the same text. There will be a loss of information from the original text. We also see that translating from the representational to the informational has some interesting properties as well. Not only has some aspect of the original informational part been lost, but an interpretation from the entity has also crept into the information. The entity does this because they have their own, what we call, perceived reality of different things in the world. In this case, of different languages and of the original author of the text. No two entities have the same perceived reality of different things in their environment. We will see perceived reality come up again when we discuss models, an entity's interpretation of an entity, resource, or an action.

4 Concepts

The majority of Information Dynamics concepts can easily be explained through an example. In this paper, we present the view of a doctor working in a hospital who will be treating patients. This example has been simplified in order to understand the basic principles of Information Dynamics. It will be assumed that the doctor will only have one patient at a time and the patient will not leave during any of the medical services performed. However, it is entirely possible to have a much more complex series of states with the exact processes a doctor goes through with several patients. We present the example in a series of process states in figure 1.

While it may appear this finite state automaton of a doctor attending to a patient shows nothing more than several different processes, this allows us to recognize that many ways of doing things, including medical staff administration and a product assembly line, have been in practice for a very long time and will not fundamentally change. This way, we can examine the information that each state of the process contains in order to develop systems which expose information first hand and allows us to make critical decisions on what to include in the software and what kinds of extensibility of a particular part of a process needs. Also, when looking at the process itself, it allows the system designers to expose necessary information to all states, no just the process state at hand.

4.1 Entities and Mechanisms

Any autonomous being or machine which imposes information to the system and/or receives (or senses) information from the system can be considered an entity. Living entities which can do both include the medical doctor, the patient, and any other humans that may enter the picture including nurses and patient's relatives and friends. However, machines also impose and receive information. For example, in the treatment phase, machines may administer certain fluids, food, and drugs to a patient. Doctors, either themselves or doctor's assistants, would setup the machine to give the correct amounts of an administrable item and then setup proper operating conditions and vital signs. After the doctor has setup the machine, it becomes completely autonomous. By itself, it will administer correct doses of what has been specified to the patient and then monitor patient's life signs through its sensor mechanisms. If the patient's life signs stay in the proper range, then everything has gone according to plan. However, if the patient's life signs stray out of the range or if the machine malfunctions, it will impose information since it will either have a lack of information from the malfunction or it will inform the medical staff of a patient's life signs have strayed away from the correct operating conditions.

We can say even inanimate objects both impose and sense information from the system. The robe which a doctor wears may have a certain color. In this sense, it has imposed the information that it has a certain color. However, the color would not appear in the absence of light. In this sense, it would appear that the robe has taken information from the world. However, one can argue that the light that gives us the information about the robe. But, even then, the light would not be able to reflect the color without the presence of the robe. It should be considered unclear whether or not inanimate objects both impose and sense information from the system. This paper will not argue any further as to whether or not an inanimate object, such as a robe, can impose and sense information from the system.

Without entities, there would exist no system. They act on information contained within a particular system. To do this, they utilize their mechanisms of imposing information and sensing information. They impose information by communication or performing an action (actions will be discussed in detail later). Sensing information can be likened to utilizing the five senses. These mechanisms of hearing, seeing, touching, feeling, and smelling all give us information about particular things.

4.2 Resources, Time, and Location

In the system, a numerous amount of resources can exist at any one time. If a medical doctor wants to evaluate what kinds of treatment options they may have, he or she would have to look at the resources they have available to them at the appropriate time. For instance, if a patient needed to go into surgery, they would have to make sure that an operating room would be available, the appropriate staff be on hand, and the right tools should be there.

The resources just mentioned have two intrinsic properties to them, time and location. Everything must be available at a certain time and be at the proper location. If something cannot be provided at the correct time or at the right location, then it, the information or resources, can almost be considered useless. We will see time and location taken into account more in-depth when looking at planning and executing information actions.

4.3 Models, Perceived Reality, and Abstraction

In the section concerning information and its representation, we briefly introduced the notion of perceived reality, the interpretation of one entity of another entity, group of entities, actions, resources, or a system. Perceived reality differs from one entity to another entity. To capture these facts, perceived realities can be thought of models. Models should be taken as a set of facts and relationships that one entity has of another entity.

To explain this better, let us take a look at the medical doctor example again for each stage. When the doctor waits for a patient to arrive, he or she would have models of the hospital (who works there, what capabilities they have, what instrumentation they can provide, etc.) and a model of several medical conditions. The modeling of these conditions could include the symptoms for these conditions, what the outlook of a patient is given a certain condition, and what possible treatment can be given. When a patient arrives, the doctor would use all of the models he or she has to begin the initial diagnosis.

During the initial diagnosis a doctor will look for symptoms and ask the patient or others certain questions. As new answers come about, the doctor will abstract certain parts of their entire model of the medical world in order to frame it in a way that fits what kinds of problems the patient may have. After diagnosing the patient for the first time, if their ailments can be cured, the doctor will formulate treatment.

Treatment can only be administered if the proper resources can be provided to the doctor at the right place and at the right time. For example, if the patient needs an emergency heart transplant, then there had better be a heart available within a specified range and time because the heart must be transferred in a timely manner after being obtained. After a treatment has been decided and performed, the doctor will then evaluate the patient to see if the patient is recovering and cured or if the patient has become worse or the treatment has had no effect. This evaluation will be done with respect to a model of how the patient should be under normal circumstances. Should the patient becomes worse or if the treatment has no effect, the doctor must then reformulate his original diagnosis. This reformulation will take into account the entire space of medical problems as well as information from the previous model he or she constructed in order to create the model for treatment. The result of the reformulation may be that the patient cannot be cured or that another treatment may be possible. The new treatment would be derived from the altered model the doctor had created.

In order to evaluate models and their usefulness, we should consider that information contained within models will constantly be fused together in order to infer new information. When a doctor tries to make a diagnosis, he or she would take several facts together. This could include temperature, physical signs, and blood tests. Individually, these may not mean anything in particular. Together, they could comprise the actual list of what an ailment contains. This fusion of information can be considered inferring new information which could be useful later, such as in figuring out the appropriate treatment.

4.4 Information Movement and Actions

In the medical environment, especially a hospital, information must be moved around at a constant rate. When examining a patient for the first time, blood tests and DNA tests are often done in different departments. The results of these tests must be given to the attending doctor fairly rapidly since especially if the patient may be in a critical state. Movement of such information must be addressed on how to do it and when to do it.

Figuring out what actions must be taken when a patient arrives and while under care must be done with careful consideration. Planning such actions will always be done with respect to the doctor's models and perceived reality. If a patient shows symptoms which require a blood test, the doctor must plan how much blood he or she must take from the patient and when they need the results by. Using this timing, they can plan their course of action over a period of time. Of course, the doctor cannot take into account unforeseeable circumstances such as a long wait for completing a blood test. Failure or troubles should never be considered an exception, but commonplace in the Information Dynamics framework. We will shortly see that planning and executing actions will be greatly affected by how much we value information.

4.5 Trust, Value of Information, and Information Uncertainty

Each piece of information we receive have different values depending on the time, location, entities, and relationships. Suppose a patient has been administered a drug and then has an allergic reaction to that drug. The doctor knows this now and then takes recourse action. The next minute the patient's history comes in and indicates that the patient has allergic reactions to certain drugs. This new information that has come in did not come at the right time. This information now has little value since the doctor knows it. However, if the information came before the drug had been administered, the value of this piece of information greatly increases.

Now consider location. If the piece of information that indicated the patient's allergies had been faxed to the wrong location in the hospital, then that too can be considered unimportant information to a particular group that received that information, even if it appeared next door to the patient. Until it goes to the right location, information can be considered essentially useless.

Regarding the value of information, we may also consider how much an entity trusts a particular piece of information. If a patient were to be diagnosed by one doctor, that diagnosis may not be considered a very accurate piece of information since the doctor could have been a pediatrician trying to aid a triage center. If the same patient were to have been diagnosed by an actual emergency room doctor, the level of trust in that information may be different than that of the pediatrician's, regardless of who had given the right diagnosis. Issues do arise when two people of the same status give conflicting statements, how would this be resolved? The person attempting to interpret both people would use their model of the other entities and the world in order to try to resolve it.

An interesting discussion on the value of information follows: what if an entity has a piece of information which does not fit in one of their models? What is the value of information and what do we trust about it? For one thing, the world imposes so much information to entities they cannot possibly retain or use all of it. However, just because an entity does not store or utilize a piece of information does not mean it should be considered worthless. The piece of information may not fit a model now, but what about later? What happens when relationships can be found which links this piece of information we do not know about now later? It is our belief that information must be able to be stored and accessed at anytime, because we do not know when information becomes useful, even if its usage cannot be determined immediately.

Entities can never have one hundred percent trust or full value of information. They can never have a completely accurate model of another entity, resource, or system. Their perceived realities will dictate how their models will turn out of others based on the currently known relationships and values. Due to this fact, entities can never be one hundred percent sure what an action's outcome will be. We consider this fact information uncertainty. We can never be absolutely certain about anything in the system.

5 Related Work

We present only a small portion of related work here since Information Dynamics spans across several fields within computer science and, as we have seen earlier, other disciplines such as engineering and philosophy. Understanding other research, ideas, and practices in all disciplines, not just those in computer science, must be emphasized as the ideas contained in Information Dynamics span a wide variety of applications and architectures. We hope the references provided here will give a substantial starting point to learning about the various fields Information Dynamics spans. Work done on understanding the basic nature of the term information has already been discussed in section 2.

Agrawala, Larsen, and Szajda [2] presented the initial framework of Information Dynamics. Applications utilizing the ideas contained within the original paper included link-state routing [6], autonomous agent coordination [12], and digital library interoperability [13]. In this Information Dynamics paper builds on the original key points and discusses newer ideas of how to design, build, and maintain an information-centric system. The previous applications can easily be mapped onto the new framework with the additional benefits it provides with the new ideas. Zhuge [19] presents several features that we present here, such as the requirements of space, time, structure, relations, worth, inference, etc. He presents this in context to work in the e-Science Knowledge Grid Environment, but does not go into too much detail about the requirements.

Classical and modern software engineering techniques [16] have been around for years in order to understand the correct way of building software systems. However, many of these techniques can be considered process-oriented and may or may not integrate the important information at hand very well. McComb [14] proposes that semantics of the processes should be considered when developing the system well before actually writing the code for it. We should take semantics to mean developing information and relationships of the system, in McComb's case, a business system.

Representation has been an issue with database systems for decades [9]. The majority of database management systems would eventually be placed on top of a relational database. Today, we see more and more data being structured in semi-structured data [1] and much research exists that work on it today.

Programming languages and software engineering have had model checking techniques available to them [5]. In order to verify that a program behaves correctly, the program is abstracted into a series of states with conditions that must be met at each state. This would seem appropriate for Information Dynamics, but model checking does not account for what happens when an error occurs. Errors should not be considered the exception, but occurring frequently.

Trust has been explored in many areas. In particular, trust in social networks has been explored in detail [8] [3]. Many times, trust has been assigned to entities for use in multiple domains. However, trust should not be assumed this way since an entity may trust another entity for one thing, but not the other. An example has been presented earlier concerning a pediatrician and an emergency room doctor treating a patient in a triage.

6 Conclusion

We have presented our vision for the development and the maintenance of computer and software systems in this paper, called Information Dynamics. This information-centric approach spans decades of research in all fields that must be fused together in order to understand the needs of a particular system. We will continue looking into Information Dynamics, in theory and in practice.

References

[1] Serge Abiteboul. Querying semi-structured data. In *Proceedings of the 6th International Conference on Database Theory (ICDT 1997)*, pages 1–18, 1997.

[2] Ashok K. Agrawala, Ronald L. Larsen, and Douglas Szajda. Information dynamics: An information-centric approach to system design. Technical Report CS-TR-4108, University of Maryland, College Park, 2000.

[3] P. Oscar Boykin and Vwani P. Roychowdhury. Leveraging social networks to fight spam. *IEEE Computer*, 38(4):61–68, April 2005.

[4] Andrezj Chmielecki. What is information? In *Twentieth World Congress of Philosophy*, August 1998.

[5] Edmund Clarke, Orna Grumberg, and Doron Peled. *Model Checking*. MIT Press, 2000.

[6] Hyeonsang Eom, Ashok K. Agrawala, Sam H. Noh, and A. Udaya Shankar. Information dynamics applied to link-state routing. Technical Report CS-TR-4297, University of Maryland, College Park, 2001.

[7] Luciano Floridi. What is the philosophy of information? *Metaphilosophy*, 33(1–2), January 2002.

[8] Jennifer Golbeck, Bijan Parsia, and James Handler. Trust networks on the semantic web. In *Cooperative Information Agents*, 2003.

[9] Joseph M. Hellerstein and Michael Stonebraker. *What Comes Around, Goes Around*. MIT Press, 2005.

[10] David Israel and John Perry. *What is Information?*, pages 1–19. University of British Columbia Press, 1990.

[11] David Israel and John Perry. *Information and Architecture*, volume 2, pages 147–160. 1991.

[12] Andrez Kochut, Ashok K. Agrawala, Ronald L. Larsen, and A. Udaya Shankar. Ad (attacker defender) game. Technical Report CS-TR-4265, University of Maryland, College Park, 2001.

[13] Ronald L. Larsen. Information dynamics: An information-centric approach to digital library interoperability. In *Global Digital Library Development in the New Millennium (NIT 2001)*, pages 137–144, 2001.

[14] Dave McComb. *Semantics in Business Systems*. The Savvy Manager's Guide. Morgan Kaufmann, 2004.

[15] Gerry McGovern. What is information? New Thinking E-Mail Newsletter, August 1998. http://www.gerrymcgovern.com/nt/1998/nt_1998_08_03_what_is_information.htm.

[16] James McGovern, Scott W. Ambler, Michael E. Stevens, James Linn, Vikas Sharan, and Elias K. Jo. *A Practical Guide to Enterprise Architecture*. The Coad Series. Prentice Hall, 2004.

[17] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423,623–656, July and October 1948.

[18] Christopher Gordom Timpson. *Quantum Information Theory and the Foundation of Quantum Mechanics*. PhD thesis, The Queen's College, 2004.

[19] Hai Zhuge. The future interconnection environment. *IEEE Computer*, 38(4):27–33, April 2005.